A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

The Principal Replicated Factors Discovered in Objective Personality Tests ... 337
Raymond B. Cattell

The Level of Aspiration of Schizophrenic and Normal Subjects ... 343
Kenneth E. Zerf

Attitude Formation: The Development of a Color-Preference Response Through Mediated Generalization ... 351
Bernice S. Eiduson

The Relationship Between Symbolic and Manifest Projections of Sexuality with Some Incidental Correlates ... 363
Russell A. Clark and Maude Von Sneidar

Test Anxiety and Blame Assignment in a Failure Situation ... 375
John Doris and Seymour Sarason

Einstellung Effect and Goal-Seeking Behavior ... 379
Norman L. Harvey

Problem-Solving Rigidity and Decision Time ... 385
David Lindon

Rigidity and Reactive Inhibition ... 395
Robert B. Malmo and Harvey Vallerand

Responses to Repeated Frustrations of Young Children Differentiated According to Need Area ... 449
Nancy Bayley Gist and Boyd R. Candler

Continued on inside front cover

Volume 50 MAY 1955 Number 3

CONTENTS—*Continued*

ARTICLES PAGE

The Influence of Stress on the Perception of Incongruity 351
Charles D. Spielberger

A Proposed Method for Increasing Accuracy of Judges' Classification of Verbal Material 357
S. Rosenberg and B. I. Langer

A Field Experiment on Rumor Transmission and Distortion 363
Stanley Schachter and Harvey Burdick

Changes in Goal-Object Preference as a Function of Expectancy for Social Reinforcement 372
David E. Hunt

Learning and Retention of Words of Sexual and Nonsexual Meaning 378
Maurice G. Kott

Thinking Deficit in Schizophrenia and Changes with Arousal 383
Rita Senf, Paul E. Huston, and Bertram D. Cohen

Phonetic Symbolism in Natural Languages 388
Roger W. Brown, Abraham H. Black, and Arnold E. Horowitz

CASE REPORT

Somatic Therapy as Discussed by Psychotic Patients 396
Libby Fich and Leslie Nawas

CRITIQUE AND NOTES

Early Language Behavior in Negro Children and the Testing of Intelligence 401
Beauford A. Passamanick and Hilda Knobloch

Respondent Reactions to Ambiguous Items in an Attitude Scale 402
Robert O. Blood, Jr.

The Selection of Stimulus Words for Value-Duration Threshold Experiments 403
David Rosenthal

Regression Among Girls 405
William W. Wattenberg and Frank Saunders

REVIEWS 407

NEW BOOKS RECEIVED 415

Volume 50 MAY 1955 Number 3

The Journal of Abnormal and Social Psychology

Editor: J. McV. Hunt, University of Illinois
Associate Editor: Arthur Jenness, Williams College
Book Review Editor: Stanley G. Estes, Harvard University
Managing Editor: Lorraine Bouthilet

CONSULTING EDITORS

Gordon W. Allport
Harvard University

Donald T. Campbell
Northwestern University

Hadley Cantril
Princeton University

Charles N. Cofer
University of Maryland

Lee J. Cronbach
University of Illinois

John Dollard
Yale University

Leonard W. Doob
Yale University

Charles W. Eriksen
Johns Hopkins University

I. E. Farber
State University of Iowa

Leon Festinger
University of Minnesota

Charles M. Harsh
U.S. Navy Electronics Laboratory, San Diego

L. S. Kogan
Community Service Society

W. Ernst Kris
Yale University School of Medicine

Donald W. MacKinnon
University of California

Quinn McNemar
Stanford University

O. Hobart Mowrer
University of Illinois

Lawrence I. O'Kelly
University of Illinois

Charles E. Osgood
University of Illinois

Donald R. Peterson
University of Illinois

Saul Rosenzweig
Washington University

Julian B. Rotter
Ohio State University

Harold Seashore
Psychological Corporation

David Shakow
National Institute of Mental Health

Richard L. Solomon
Harvard University

Ross Stagner
University of Illinois

Ivan Steiner
University of Illinois

Donald Thistlethwaite
University of Illinois

Elizabeth Collins Wesman
White Plains, N. Y.

Robert W. White
Harvard University

Charles Wrigley
University of Illinois

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

Entered as second-class matter February 15, 1954 at the Post Office at Baltimore, Maryland under the Act of February 28, 1925. Acceptance for mailing at special rate of postage provided for in Section 34.40 (e), P. L. and R. Additional entry at Washington, D.C.

Copyright 1955 by the American Psychological Association, Inc.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 816 pages. The subscription price per year is $10.50; foreign, $11.00. Single issues are $2.00.

Original manuscripts should be sent to the Editor-elect:

DR. M. BREWSTER SMITH
Social Science Research Council
230 Park Avenue
New York 17, N. Y.

Correspondence about articles to be published in 1955 should be addressed to:

DR. J. MCV. HUNT
Department of Psychology
University of Illinois
Urbana, Ill.

Books for review should be sent to:

DR. STANLEY G. ESTES
Psychological Clinic
64 Plympton Street
Cambridge 38, Mass.

Fifty free offprints are supplied to contributors of articles. For book reviews, five copies of the JOURNAL are supplied gratis if requested. Authors of prior publication articles receive no gratis offprints or journals.

Subscriptions and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street N.W.
Washington 6, D. C.

Printed and mailed by:

WAVERLY PRESS, INC.
Baltimore 2, Maryland

Changes of address must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

THE PRINCIPAL REPLICATED FACTORS DISCOVERED IN OBJECTIVE PERSONALITY TESTS

RAYMOND B. CATTELL

Laboratory of Personality Assessment and Group Behavior, University of Illinois

The Logic of Factor Measurement in Personality Research

That advance toward establishing scientific laws in personality study depends on the inauguration of precise and meaningful personality *measurement* is today a widely recognized principle. This principle has recently been stated and examined, for example, in such diverse approaches to personality research as those of Eysenck (31) and Hunt and Kogan (44) as well as in an earlier systematic exposition (7). In any case, mature researchers have long been forced to perceive that crucial enterprises in theory, as well as the practical efficiencies of applied psychology, hinge on such a development of structural measurement.

Fortunately, the rising generation has become familiar with the concepts and techniques for basing research into personality structure and dynamics upon multivariate methods, notably factor analysis and its various developments, capable of yielding previously unknown patterns and concepts. Without objective and analytically powerful techniques of this kind, personality research must depend either on clinical concepts, more subjectively derived and seldom expressible in operational measurement to everyone's satisfaction, or on arbitrary, atomistic "scales" or patent tests, which can be at least as numerous as psychologists!

The logic of the factor analytic approach (13) resides in the statements that (*a*) functionally unitary traits or processes are to be discovered by observation of covariation, among persons, or in one person over time, with or without controlled experiment, (*b*) concepts are better operationally defined by, and based upon, a *pattern* of measurements rather than by a single measurement, as in classical univariate experiment,[1] and, (*c*) a functionally unitary trait or process should nevertheless not be considered established by a pattern in a single factor analytic research, but must reappear consistently and persistently in independently rotated studies.

To achieve this last condition, research must be pursued with common variables, continuity, and an architectonic plan. The technical, organizational, and social obstacles to achieving these standards are doubtless responsible for the unfortunate fact that this demonstrably rewarding and indispensable line of research has received contributions from extraordinarily few of the many psychologists interested in advancing personality research, and has fallen short of closure. Moreover, as French (35) has pointed out, a sad proportion of even the factor analytic contributions to personality measurement actually fail, on technical grounds, to be finally of any use. This should be borne in mind if the reader is not wrongly to assume that our concentration on the work of so few laboratories betokens disinclination to make a sufficient survey.

Our purpose has been to get the best possible indication from all existing research as to the *number and nature of factors in personality measurement that show reasonable stability from study to study*, and which could next be made the conceptual basis for a definitive checking research. Interpretation and theoretical discussion must, for reasons of space, be largely deferred (15).

Data, Methods, and Concepts in the Survey

Since 1945, when material was collated for the present writer's integration of personality

[1] Even at the present day it is not unusual to find a single alleged concept; e.g., anxiety, rigidity, projection, the hunger drive, the ego, employed in the same theoretical framework by investigators but operationally "defined" by variables as numerous as the different investigators. The converse ambiguity, in which agreement on a similarly-titled "scale measurement," adhered to in different researches, nevertheless hides a confusion of several independent factors in the "scale" is also common. A recent instance is the work relating "anxiety" to conditioning, in which no attempt was first made to determine the number and nature of factors in anxiety responses by factor analysis. Instead, the traditional semantic error was made of assuming that one word means one thing.

measurement data (7), no further survey has appeared until 1953 (35). A survey of *ability* factors has been made by Guilford (37) and by the committee working with French (34), and the latter has recently surveyed personality factors in ratings, questionnaires, and objective tests (35). The present study is independent of the last-named, and the weightiest part of its evidence is additional to any available at the time of French's survey.

By objective tests are meant those in which introspective, deliberately self-assessing responses are not used, and in which behavior is measured in ways such that different experimenters (*Es*) will assign the same score to a given subject's (*S*'s) response. They include miniature-situational, misperceptive (projective), stylistic and physiological devices, but also various novel forms, equally important, for which no name has yet become current. By personality factors we mean temperamental and dynamic factors, as defined and distinguished from abilities by basic criteria set out elsewhere (7). However, in practice the distinction is not always easy, and, to be safe, we have left the lists open to the possibility that some factors in perceptual ability tests are personality factors. Although the factors found in specifically dynamic and learning data (interests, attitudes, memorizing, symptoms, etc., objectively measured) must eventually be considered in close relation to general personality factors, we have set them aside as too few, and as only once confirmed (23). Were such confirmed factors available, it might be hard to draw the line between such dispositions or drive strengths, e.g., timidity and amorousness (which are clearly "dynamic"), on the one hand, and a "dynamic integration" factor as a *general personality dimension* on the other. But drives, self-sentiment structures, etc. can be roughly distinguished, for the present, by their specificity of object attachment, from "general personality" factors (largely temperamental) pervading all behavior, and we shall leave them to a distinct survey a year or two hence.

Space need not be given here to describing the factor analytic method (2, 13, 41, 62); nor does it permit us to do more than merely state here the following conditions which we believe have been shown elsewhere to be necessary for unique, psychologically meaningful solutions:

1. The factors must be rotated for simple structure, since unrotated centroid or principal component solutions or other resolutions based on merely mathematical symmetries have no reason to be invariant from research to research. Our standard procedure implies oblique factors and "blind" decision on simple structure, without reference to favored hypotheses.

2. Marker or pilot variables must be carried—at least two to any supposed factor—from earlier studies, so that new factors may be located in relation to the old. It is not enough to have "psychologically similar" tests: exact "dovetailing" of studies requires *identical* tests.

3. Regard must be paid to the "density" of variables (13) and their due sampling according to the personality sphere concept (7) in order to avoid mistaking what is psychologically a specific for a general personality factor, or a second-order for a first-order factor. Besides, this comprehensiveness of sampling provides sufficient "hyperplane stuff" to locate good simple structure (13).

4. The usual, routine statistical requirements must be met, e.g., an adequate sample, iteration of communality estimates, extraction of the correct and sufficient number of factors (without which rotation is distorted), sufficient variables to define the number of factors, known reliabilities for variables, and "blind" rotations for simple structure, carried to an unimprovable "plateau" number of "total variables in the hyperplanes."

These conditions have frequently been met in the study of *abilities*, e.g., the extensive studies of Thurstone (60, 61, 63), Guilford, *et al.* (37, 38, 39). However, in *objective* tests of *personality* they are at present met by only about a dozen published studies, seven of which are from the Laboratory of Personality Assessment at the University of Illinois. Consequently, these must be the central core on which we organize our discussion. The published studies of Eysenck are included, despite being somewhat defective for the purposes of the present integration (since they extract too few factors and are not rotated for simple structure,[3] unlike his further

[3] Thurstone's factor analytic work is now proceeding to objective personality tests, but results are not yet

studies in progress) because of their criterion association.

Although several P-technique (13) factors seem to retain recognizably the same form as R-technique factors, the present survey, with its necessary standards of population matching, will be confined to the R-technique domain, the P-technique findings being deferred to a separate collation (16a).

DESIGN IN THE SEVEN STUDIES

Since the seven core studies about which others in the present paper will be organized constitute a single strategic research, organized as such from the beginning over a ten-year period, a section must be devoted to describing the objectives and rationale of that enterprise.

At the commencement, for historical continuity, marker tests were included for some twenty-two promising factors in past research, chiefly as summarized in the 1946 survey (7) and including Spearman's "Fluency of Association" (57), Thurstone's perceptual factors (61), Perseveration-Rigidity (27, 51), Dynamic Momentum (7, 54, 55, 56) (sometimes called "satiation"), Speed (47), Tempo (7), Endurance (54, 55, 56, 59), Will Factor (1, 7, 29, 52), and Honesty (1, 7). Markers for these were taken principally from Brogden (1), Crutcher (29), Gibb (36), Hartshorne and May (40), O'Niel (53), Rethlingshafer (54), Ryans (55), Thornton (59), and others. An attempt was made also to include, through personal communications—notably with Eysenck, O'Niel, and Thurstone—whatever markers were appearing in a current work. Finally, there were included some specific tests with personality predictive power, e.g., the C.M.S. (6), Mira's Myokinesis (50), suggestibility (7), aspiration discrepancy, color-form ratio, ergograph effort curve, etc., as described elsewhere (15).

Beyond this linkage with the past the whole research effort was concentrated on the creation of *new* tests for the new personality concepts emerging from factor analytic research. The hypotheses of test design were finally made explicit for all our tests, though they may have begun more intuitively. When they began explicitly they were formed around the conceptions of personality dimensions (e.g., surgency, ego strength, schizothymia) derived from the earlier studies on factors in *behavior ratings and questionnaires*. Indeed, it will be recalled that the larger plan (5, 7) within which the ten-year objective test plan was designed called for the *initial* delineation of factors in terms of criterion, "everyday-life record" behavior, including behavior ratings, and that the objective test design began as soon as these earlier factors were stabilized (7). This spiral of hypothesis formation and factor experiment apparently needs to be pointed out because it is sometimes said, even by factor analysts, that factor analysis is not a hypothetico-deductive method. It is true that it *can* start without hypotheses and create them in the factors yielded, in a way not found in univariate experiment (13). But, it can also *check* hypotheses in a peculiarly thorough manner, and in the present researches this methodological cycle of creating and checking has gone from use of ratings and intuitive, clinical hypotheses to the final design of objective tests by clarified hypotheses. Thus, to a factor analyst, the computational work here has been only a matter of keeping standards of everyday workmanship, and the real fulcrum of effort has been the designing of tests on clear psychological principles.

Some 330 test variables (but rather fewer test situations), consisting first of about 60 borrowed "historical markers" and then of 270 newly devised measures, have thus been tried out in the past ten years. In the methodological cycle—continued through seven distinct major factorizations and certain *ad hoc* "patching" studies (7, 8, 12, 19, 20, 21, 22, 24, 26, 27)—these original tests have sometimes been improved to give higher loadings as the factors concerned converged on more definite meanings, permitting desirable modifications to be perceived.

This possibility of improving test validities creates a dilemma in regard to "marking" factors, for the firmest basis of matching is provided by carrying a continuous track of

published except for specifically perceptual tests, included below. Burt's extensive personality factorizations do not deal with objective tests. Sundry factorizations of specific objective tests, both published and in theses, are on too narrow a range of variables, e.g., are confined to "rigidity" or "oscillation" or autonomic variables, or parts of a particular test. We have included Eysenck's factors, despite the above objections, because, even without any rotation, the grosser factors at least retain recognizable form.

absolutely unchanging tests through all factorizations. However, improvements can be made on *different* tests in a factor at alternative cycles, provided sufficient absolutely *identical* tests be carried between immediately successive researches. The latter "minimum necessary overlap" compromise is defective in that factor identification A = C has to be established by A = B and B = C, which gives more chance for error. A flexible compromise of the two was actually adopted: researches were cemented by a continuous track of immutable variables through all matrices, while some newly discovered or improved high-loaded tests were also carried from the point of discovery to adjacent or later researches. Incidentally, it will be observed in the subsequent tables that the original markers of the former kind tend to sink from loading prominence as the researches proceed. This is partly the usual regression from a first to a second sample and partly a contrast effect from their overshadowing by the newly devised tests. Close examination will show that they stay significant in the factor to which they belong, relative to any other.

The next methodological point requires a reminder that our objective has been simply to locate the primary personality factors in a widely conceived range of behavior measures, and to permit their reproduction with acceptable invariance of pattern by others. We are *not* aiming at the unreal goal of getting immediate loadings high enough for complete interpretation or for estimates of individuals' factor scores beyond the modest reliability of about 0.5 to 0.8. Consequently, we have not set up test lengths to give the usually desired reliability of individual tests nor have we item analyzed tests in most cases. The short and therefore less reliable test is absolutely demanded by the aim of including *a large number* of tests in a single matrix (and within the subjects' time). And, this large number—about 48 to 128 in the batteries of the present studies—is indispensable for (*a*) including old markers and yet breaking new ground, (*b*) obtaining a good hyperplane for any one factor, (*c*) "blocking out the whole picture" of primaries, and (*d*) avoiding as far as possible "untraced variance" for any variable, such as brings uncertainty in factor identification.

In this field, perspective is everything; the notion of fitting together a host of narrowly conceived but highly exact one-or-two-factor researches is illusory (13). There is a time for approximation—for the acceptance of calculated risks—and it comes at this stage of research. Low reliabilities do not systematically distort the pattern of loadings of any variable. One can still see in which factors it is highly involved and in which little, and estimates of the true value can be obtained by correction for attenuation. As in a perceptual gestalt, the *shape* transcends the condition of the elements, and it has been empirically shown that tests of low reliability retain their factor pattern (23) with remarkable consistency from study to study. Besides, it is a waste of research time to bring all of some hundreds of new test devices to the finished reliability of item-analyzed tests when only a minority of them will turn out to have any personality factor relevance. Nevertheless, this careful strategy has been confused by the compulsive with inaccuracy. But, if it provides a perspective of the primary factor patterns, so that more restricted researches can confidently go to work on individual factors, to raise the reliability of tests and introduce still more valid tests of those factors, it will have abolished a bottleneck which has long held up a wide area of possible quantitative reasearch in personality.

MATCHING AND EVALUATION PROCEDURES

The breadth of scientific meaning to be attached to the subsequently matched factors depends on initial decisions as to which features in the seven studies should be deliberately controlled or varied before matching. Thus, since computing method should be irrelevant to the nature of psychological unities, we deliberately used simple centroid on the first study, the grouping method on the second, the multigroup method on others, and have obtained similar results using principal components on a still later study. Rotations, similarly, were by different methods and people, in ignorance of other findings and working with the single objective of maximizing the number of variables in the total hyperplanes, plotted until a plateau for "total hyperplane content" was reached.

On the other hand it seemed necessary, until we should get to know how "hardy" patterns are in response to changes of popula-

tion, to have our samples in the main from essentially the same population—actually, young men and women of rather more than average education. Nevertheless, two distinct variants were also tried, namely 11-year-old children and hospitalized psychotics, and some of the deviations of pattern in the tables must be evaluated in terms of this change of population.

For economy of time and space we have used the following system in referring to researches, at the head of the columns of factor loadings recorded in all subsequent tables: (*a*) we have referred to researches by the initial of the researcher, C for the present writer, E for Eysenck (31), T for Thurstone (60, 61), Th for Thornton (59), Cn for Connor (28), Br for Brogden (1), T-R for Thornton and Ryans (55, 59), roman numerals as used in the Index in reference 7 for "combined" and "composite" patterns already averaged in the present writer's earlier survey (7). The numbers attached to these letters describe the successive researches of the given investigator, e.g., E1, E2, E3 refer to the three successive E *studies* in the bibliography (31, 32, 33). (*b*) We have indicated above the letters the size of the experimental sample. This also includes a reminder of which samples are *not* from normal adults by the addition to the number of "a" for abnormal and "c" for children. (*c*) We have referred to tests by numbers of the *Master Index* (kept in the Encyclopedia of Tests in this laboratory and not yet published as a whole) into which the specific matrix numbers, in particular published researches, have always been translated in the original publications, and (*d*) we have referred to factors by a Universal Index, as proposed in detail elsewhere (14) which continues from U.I. 15, the last of the factors agreed upon by the committee on ability factors (34). Thus, the best established twelve personality factors run from U.I. 16 through 27 and the less certain factors (not listed here) from U.I. 28 through U.I. 33. The separate articles (8, 12, 19, 20, 21, 22) must at present be consulted for description of the tests, which are here given by bare title and M.I. number; but a work in preparation (15) and the actual Handbook for these tests (16) will give details of construction, administration, etc.

Now the first task before us is to develop objective means of matching factors, on the provided *common* variables, from study to study. For psychologists must at this stage of research eschew the conceit that matching can be established by mere "insight" into psychological meaning. When two factors, each in an independent study, appear similar, by reason of having substantially the same variables highly loaded and in the hyperplane, the similarity may be more exactly examined by (*a*) the correlation coefficient, r, or pattern similarity coefficient, r_p (10), applied to the loadings to express the similarity of pattern profile, or, (*b*) the salient variables similarity index, s, to express the likelihood of encountering the given number of *salient* variables found in common by chance. (By salient variables we mean those which the psychologist commonly considers as characterizing the factor.) Both instruments were used here, any peculiarities of action of either being partly avoided by their consensus. The latter instrument, especially developed for the factor matching problem (9, 17), has the following formula:

$$S = 1 - \sum_{x=0}^{x=c-1} \left(\sum_{y=0}^{y=g-x} \frac{\binom{g}{x}}{2^g \binom{n}{g}} \cdot \left[2^{g-x-y} \cdot \binom{n-g}{g-x-y} \cdot \binom{g-x}{y} \right] \right)$$

where g is the number of salient variables in (either of) the two studies; n is the total number of variables common to the two studies; c is the number of variables that prove common to the two sets of salients (common in nature and in sign), x takes in succession values from 0 to $c - 1$ and y takes in succession values from 0 to $g - x$.

A systematic matching procedure has been carried out in our own seven studies by pairing every factor with every other, but this does not mean that all coefficients have been worked out, for in a majority of instances one can see at a glance that the factors are different. In most cases a single match stands out, and even if the r or s (inverted) is not large it is usually significantly larger than for any competitor. But in a few cases, mentioned below, two possible matches in series 2 for factor A in series 1 had to be closely compared—a

dilemma which we believe to be specifically associated with the phenomenon of "cooperative factors" (13).

The above indices may be illustrated on familiar ground, so that the reader will have more "feeling" for these new, goodness-of-matching values in the ensuing personality factor context. For this purpose we may take the general ability factor. Although our research was casting its nets for newer, unknown patterns, this ubiquitous pattern seemed inevitably to be fished up in each and every study. Naturally, it is not as well marked as if we were specifically choosing tests to define its pattern, but we always carried one marker for it and for the rest it picked up those personality tests which clearly involved intelligence.

In Table 1 which sets out the general ability or "g" factor as it occurs in nine studies, we have followed the procedure used in the remaining twelve factor tables which may therefore best be immediately described as follows. (*a*) Initials and numbers are used at column heads, as described above, to define the research by its author and its position in his series of researches. (*b*) Size of sample is placed right above this symbol. (*c*) "Factor in Series," above *b*, gives the number of the factor in the series of factors *as reported in the research in question*. A minus sign after it shows that it has been reflected relative to its original table. (*d*) In the rows (the variables), the Master Index number, already described above, is placed on the left of each variable. (*e*) The title of the test is adjusted by epithets: high, low, good, poor, etc. as dictated by the sign of the loading (or by the majority of signs where there is disagreement), i.e., the title has *already* been made to agree with the sign. (*f*) In the upper part of the table are listed the variables used in *at least* two studies, while those in the annex below the line are variables which have been used only once, but have achieved such substantial loadings that they must be presented as probably part of the factor pattern.

A tolerable agreement of signs and magnitudes is evident to direct inspection alone in Table 1. (Thurstone's factor is, of course, his

TABLE 1

FACTOR 1. PATTERN OF GENERAL ABILITY OR INTELLIGENCE

Master Index No.	Title	Factor Identification								
	Factor in Series:	6	2	3	1	2	8	2	9	1
	Sample:	370	370	100	100a	500	250	184c	—	240
	Research:	C1	C2	C3	C4	C5	C6	C7	S1	T1
11	High combined verbal and numerical ability	42	88	40	57					
26	High ability to solve riddles	27		47	50					
201	High ability on number series					54	24			50
199	High score on computation					50	64	55	18	08
167,50	Good immediate memory for attitude statements	18	54	12	48	03	22	43	33	
2	Low motor rigidity	−11	−39	−17	−69	−35	−17			
41	Good two-hand coordination	04	41	15		25		40		
159	High accuracy of self-performance estimate			−12		−46	−61			
249	Good memory for names					36	20	54	14	04
7	Fast speed on gestalt completion	05		15		32	(−01)	71		
51	High index of carefulness		27*			52	15			
65	High logical consistency of attitudes		08	54						
274	High verbal ability (Stanford)							55		
275	High arithmetic reasoning ability							52	59	33

* In all studies from our own laboratory the loading signs have been checked back, for the purposes of this integration, from the published matrices to the original test scoring. This is necessary because it is extremely easy in the various reflections which the variables have to undergo for sign errors to occur, and errors have been found in about 1 per cent of the published signs. Where an asterisk is left in any table it means that there may be doubt about the sign insofar as the original score sheets, or their re-scaling sheets, are no longer stored.

second order general ability factor derived from the primary mental abilities [60].) To appreciate the degree of agreement the reader must remember that below "the visible part of the iceberg" is a larger mass of variables which agrees in being *unloaded* in the factors in question. In terms of a double-dichotomous table (significantly loaded versus insignificantly loaded), on two compared factors, the present list actually shows three of the possible quadrants, namely: significant-significant; significant-insignificant; and insignificant-significant, omitting the far more bulky fourth, for economy of space. Indeed the variables listed (in the upper, main table) are arrived at as follows: Any variable that has an appreciable (presumably significant) loading in any of the factors purporting to match *has* to be listed. Its loadings in the other factors are then listed on the same line, *whether of the same sign or not or whether adequate or not*, in order that the reader may see the degree of agreement achieved.

The word "appreciable" is used rather than "significant" because there is not yet agreement on the expression for significance of a loading (13), but our intention is to deal with significant loadings. From indirect and empirical evidence too discursive to discuss here (but chiefly in connection with distribution plots of rotated loadings) we judge that with populations of the size used here any loading above ±.10 is "out of the hyperplane." To remain in the region of what one is accustomed to think of as the 1 per cent level of significance we have normally taken only loadings of ±0.35 or higher, but for reasons given below we have occasionally dropped even as low as ±0.15. (The frequency distribution of recorded loadings for *all* matched factors here is: 230: ±0.35 or above; 45: ±.30 to ±.34; 42: ±.25 to ±.29; and 55: ±.20 to ±.24. The frequencies in the lower ranges increasingly include loadings merely brought in as matchings for adequate loadings in the other factors.) Where a variable in one study was simply not used in another and therefore *could not have a loading*, the loading space is left blank.

Although we have checked the matchings with *r* and *s* indices (most of which are set out elsewhere [15] for economy of tables here), these indices are only contributory pieces of evidence in a larger realm of evidence which belongs to experience in this field and which we have frankly allowed to overrule any *single* mechanical index or limit when our judgment required it. The art, or creative judgment—call it what you will—by which we have occasionally defied what the reader would require in a mechanical following of our own rules, in some of the more subtle matching issues, can be vindicated only by the verdict of further research as to whether the new variable or variables brought into the pattern by such a match is found when extraction of the pattern is repeated. Fortunately, such further research need no longer be an extensive exploration of new tests and factor space but can be a simple, restricted checking procedure on stated tests and patterns. However, some of the considerations in these judgments can be relatively briefly stated as follows.

Allowances have been made for the following. (*a*) Known differences of reliability of individual tests from research to research, due to more or less selected samples, different test rooms or administrators, etc. (*b*) Differences of mean score on the test from sample to sample, causing differences in loading magnitudes. (*c*) Differences in actual personality factor structure—loading pattern and correlation between factors—to be expected among different *population* species. For instance, easy intelligence subtests are generally more highly loaded in the "*g*" factor in dull than bright groups, while rigidity measures shift more of their variance from a specific rigidity factor to the intelligence factor (negatively) in abnormals (who have higher but narrower rigidity) than in normals. (*d*) Consideration of the mode of factor analytic computing process (multigroup method seems to give consistently larger communalities). (*e*) Particular ambiguities in factor rotation, e.g., one factor in C7 happened to have two almost equally good hyperplanes. (*f*) Where two factors, A and B, in Series 1 are rivals for a match with Factor X in Series 2, it has happened that one of them, A, already makes a very clear match with *another* factor in Series 2. In such a case even slight superiorities on some loadings for the A − X match must be discounted. (*g*) Slight test modifications.

The means of allowing for these will be clear. For example, allowance is made for *a*

TABLE 2

AGREEMENT OF PATTERNS OF LOADINGS FOR SEVEN MANIFESTATIONS OF THE GENERAL ABILITY FACTOR

Factor in Series:	6	2	3	1	2	8	2
Sample:	370	370	100	100a	500	250	184c
Research:	C1	C2	C3	C4	C5	C6	C7
A. Intercorrelations: *r*'s							
C2	+84						
C3	+63	+45					
C4	+43	+57	+19				
C5	+72	+53	+31	+47			
C6	+74	+34	+53	+33	+65		
C7	+04	+15	+34	+33	+15	+02	
B. Salient Variable Similarity Coefficients: *s*'s							
C1							
C2	.06						
C3	.01	.01					
C4	.34	.01	.04				
C5	.35	.04	.01	.52			
C6	.28	.31	.06	.01	.01		
C7	.30	.06	.01	.59	.01	.10	

above by heeding not merely the absolute loading, i.e., whether it is above 0.34, but also its loading relative to that which it shows in other factors in the same study, and the degree of consistency of loading with that of very similar tests, e.g., two forms of reaction time or closure or rigidity subtests.

From the inspection of Table 1, in the light of these considerations, we shall now turn to the more exact criterion—represented by the *r* and *s* values among all pairs of researches, which are set out in Table 2.

It should be observed first that there are sometimes decided divergencies of *r* and *s* values notably in the C1.C5 match which gives an *r* of .72 but is not significant (almost an even chance) on *s*. The differences are partly due to the relative values which these indices put on the lower loadings. A poor hyperplane fit can probably reduce *r* much more than *s*, which responds rather to the essential agreement on salient variables. On the other hand, *s* sometimes pays too much attention to a few salients and ignores wider agreements. It is unmistakably shown here that the significance is *usually* given as higher by *r*. Only 6 *r*'s fall below the .05 significance level, whereas 11 *s*'s do so. On either an *r* or an *s* test all matches are significant except C7 with C1 and C6, but there is a wide range of goodness—on *r* values extending from .02 to 0.84—though all are positive. The poorest agreement falls between the child pattern and the six adult patterns. It is encouraging that the indices shown below for the various new personality factors are in general no poorer than these values for the well-established general ability factor, and it would seem from this tryout that a mean *r* of about 0.4 (0.42 in this case), with the presently obtainable accuracies of rotation, must be considered to correspond to an identity. If we leave out the lower figures due to child and abnormal populations it would seem that at present we can measure factor patterns with a reliability of about 0.5.

Such a moderate correlation is consonant with the fairly wide variations of *absolute* loading for a given variable found in a given factor, as presently rotated, blindly and independently by individual judgment. If we take a variable most frequently repeated, *Immediate Memory* (which has two Master Index numbers, 50 and 167, because of slight administration changes), we find a highest loading of 0.54 and a lowest of 0.03, with a mean of 0.29. This is typical, for the mean between extremes of loadings (for any one variable) is 0.31, and suggests a sigma of the typical row in these tables of about .07. Such a dispersion of factor loading values about the true value will surprise no one familiar with factor analytic complexities; indeed, it would be astonishing if the range were less when we consider through how many abstracting processes and translations the figures pass from the raw scores to the final rotated loadings, each process introducing some degree of approximation or error. We have not found any other studies in the literature in which repetition with entirely independent rotations has produced better agreement than this for well-known factors; and the present agreements are accordingly regarded as reaching the level to be expected when true similarity of pattern exists, blurred by the defects of present techniques. The important thing is not the absolute level of loading, which is affected by countless specific influences, but its sign and its significance, i.e., on which side of the hyperplane it lies. Any inconsistencies of sign are pointed up in the present tables by placing the figures in parentheses. There is only one loading out of 52 in Table 1

with a wrong sign (but with a value of only −.01!). The chances of an agreement on 51 out of 52 signs, in random values, would be *extremely* small.

PATTERNS OF THE TWELVE MOST INVARIANT FACTORS

To keep in mind the samples upon which the factor patterns are matched it is necessary to refer to Table 3 which shows the number of common or marker variables arranged in the experimental design between each and every pair of studies. In general, the larger numbers tend to include the same actual variables as occur in the smaller numbers.

As indicated earlier the personality factors begin with number 16, the first 15 factors in the Universal Index (14) being ability factors. Though a formula has been suggested for distinguishing ability and personality factors (7), none is yet generally accepted and in fact we believe, as indicated below, that two or three factors which have been too automatically included in the extension of ability factors are better considered personality factors. To avoid what may be purely scholastic but time-consuming disputes it has seemed best to list ability and personality factors together in the order of their discovery and confirmation, in the same universal factor index (14).

In the present integration factors are numbered and presented in an order which combines (*a*) declining order of magnitude of variance and (*b*) declining breadth and certainty of invariance (matching criteria). The first twelve factors, which alone are fully set out here, are considered to meet the standard "established" and to be adequately stable for applied work, while six more are "indicated" and should be checked by research immediately, before applied work assumes their unitary character. Within each factor table the common variables are listed in descending order of fitness to represent the factor in terms of a combination of (*a*) mean magnitude of loading, (*b*) number of studies in which similarity of loading is confirmed, and (*c*) considerations arising from our knowledge of the reliability of the test in various administration situations. In most factors we have deliberately proceeded down to tests of quite low mean loading because the researches soon likely to follow, working intensively on

TABLE 3

NUMBER OF VARIABLES, OF COMMON MARKERS, AND OF FACTORS FOUND, IN SEVEN PRINCIPAL STUDIES

Reference Symbol for Study	Number of Variables Used	Number of Factors Found	Number of Variables Common to Two Studies (Markers)						
			C1	C2	C3	C4	C5	C6	C7
C1	48	11							
C2	64	11	28						
C3	117	17	34	27					
C4	104	18	32	27	100				
C5	127	16	20	22	46	45			
C6	64	15	15	14	26	26	52		
C7	128	16	24	18	48	47	73	43	

single factors, will need to take note of all possible tests, including those of borderline validity.[3] The reader not directly familiar with factor analytic processes and who may evaluate the agreement of figures by immediate impression should perhaps be reminded in scrutinizing the following tables (*a*) that a well-located dimension, approximately orthogonal to a set of well-established dimensions, may

[3] A note must be presented here on matters which scarcely concern the general reader but are important to anyone contemplating intensive research and close collation of the references. The following changes between the separate published researches and the present collation are to be noted. (*a*) Factors have been completed reversed in loading signs when necessary to agree with the accepted direction for the majority in a given table. This is indicated by a minus sign after the factor (head of column). (*b*) Single variables have occasionally been scored in different directions in different studies, e.g., Maturity of Opinion, Suggestibility, and Absence of Questionable Preferences were oppositely scored in the early studies. This accounts for some apparent reversals of sign of single variables from tables in the original specific study. (*c*) Additional checks have been made on loadings for those variables in the early studies which now prove to be important in factor definition. In three cases, two or three points of change in the second decimal place were found in quantity and in two cases a sign error. In one of our own variables and one of another contributor a sign error was almost certain, but could not be proved. (An arbitrary "value judgment," resulting in a reversal at time of scaling had reportedly been made by a departed assistant!) (*d*) Improvement of simple structure was undertaken two years ago on F4 in C1 and on four factors in the second study, in the light of increased rotation experience. The shifts were slight, but they alter four of the published factors in C2. The new factors may be calculated as follows: $F'_3 = F_3 + .25F_{10}$; $F'_5 = F_5 - .10F_1 + .10F_2 + .20F_3 + .10F_4 - .10F_6 + .10F_7$; $F'_7 = F_7 + .15F_6$; $F'_8 = F_8 + .40F_5$. The changes under 3 and 4 mainly effect loadings only in the second decimal place, but might be puzzling if not understood.

TABLE 4
FACTOR 16. COMPETENT ASSERTIVENESS

Master Index No.	Title	Factor in Series: / Sample: / Research:	Factor Identification 1(−) 370 C1	3(−) 370 C2	8(−) 100 C3	16(−) 100a C4	4 500 C5	4 250 C6	6 184c C7	1* 100 Br.
82	Long exploratory distance on maze			14			52	24	06	
270	Fast tempo of arm-shoulder swing						38	39	46	
117	Good ("Highbrow") taste in social and esthetic environment				26	48	47	12		
15	Excessive use of circles in C.M.S. test		(−03)		04	18	28	38	15	
268	Fast tempo tapping in "Two Hand Coordination"						38		76	
282	Many objects perceived in unstructured picture						35	34		
143	High total score on C.M.S.				40	03	40		01	
307	Fast speed of letter comparison						17	56		
21	Absence of questionable preferences in reading		26	57	31	22	18		15	31
35	Little suggestibility to authority		−16	−48	−26	−04	−08	(07)	(09)	−53
2	Low motor rigidity		−06	−02	−01	(10)	−21	−13	−10	−40
50–51	Detailed care in following instructions			05				22		40
34	Maturity of opinion		12	−21	(07)	−14	−11		(10)	−53
96	Big reduction reading speed by frustration				53				(−03)	
278,9	Fast reading tempo ("Ideational Tempo" in C1)		53				64			
125	High emphasis personal vs. institutional values				60					
99	High ratio of verbal to numerical ability					35				
309	Fast speed of line judgment						63			
237	High ratio distance on difficult to easy mazes						46			

* This has the standard No. T XII in (7).

yet "wobble" through ten or twenty degrees, and such slight changes can produce differences in loadings as large as any seen in these tables; (*b*) that he must bear in mind not only the agreement and disagreement among the salient variables set out here, but the "hidden agreement" in the *more numerous* variables not listed which coincide in being zero, as indicated above.

Universal Index 16. Competent Assertiveness. Here there happen to be very few common variables to test the matching of factors in studies C_1 and C_2. On the brief series available, factors C1: F.5- and C2: F.10- would fit almost as well as C1: F.1- and C2: F.3-. The correlation of the two loading patterns with most variables in common for checking (C5: F.4, C6: F.4) is 0.50. Factor 6 in C7 is most deviant, but this is understandable as it deals with children. Though 7 of the possible 60 loadings to be considered have inconsistent signs, the numerical values of all of these prove to be negligible. The pattern is one of fast, determined, effective action, and insuggestibility, good taste, and individuality of opinion. The present title is meant to be descriptive of speed, poise, ideational decisiveness, and self-assertion.

U.I. 17. Restraint-Timidity. The matching here is good—quite as good as for the intelligence factor—and the sign inversions are with one exception for loadings of a size that is negligible. It presents an instance wherein an "ability factor"—Thurstone's F or speed of perception (reversed here to *slowness* of perception)—agrees, except for a scarcely signifi-

TABLE 5

FACTOR 17. RESTRAINT-TIMIDITY

Master Index No.	Title	Factor Identification: Factor in Series: / Sample: / Research:	2 370 C1	4 370 C2	11 100 C3	8 100a C4	10 500 C5	12 250 C6	14 184c C7	F(−) 290 Th2
43	Large mean PGR deflection (all stimuli)		32	34	15	15			10	
21	Absence of questionable reading preferences		−47	−65	(23)	−10	−21		−13	
137	Much reduction finger maze activity by shock				42	18			31	
7	Slow speed of closure in gestalt completion		−15		−20	−19	(07)	(05)		−53
105	Tendency to perceive many threatening objects				44	40	(−01)	20	12	
72	Much slowing of reaction time by complex instructions			44				08	13	
150b	Much acceptance of "obvious" remarks				54	07			10	
5	High ratio of regular to irregular warned reaction time		07	46	18	10		08	(−01)	
10	Slowness to name object in dark adaptation		11							61
216	Low suggestibility to shock						31	57	(−03)	
93	Little effect of physical restriction on performance				(−06)	48				
120	Low ratio of inaccuracy to speed				−08*	−50				
3	Slow speed of judgment (social and spatial)		−02	(01)	−13	−16	−19	(02)	(09)	−39
27,218	Expressed willingness to play practical jokes		04						50	
51	High index of carefulness			54						
	Low peripheral span of perception									−65

* See footnote to Table 1.

cant difference on speed of social judgment (Master Index Number 3) with a temperamental pattern which is best labeled Restraint or General Inhibition. One of Brogden's (1) factors, No. 6 (Self-Control, Inhibition), *could* also fit, but a couple of common variables is not sufficient basis for so deciding. The factor which Eysenck (31, 33) has discussed as "Hysteria-Dysthymia" (and which is perhaps an incompletely rotated mixture of our F and H factors of Surgency and Adventurous Cyclothymia [11, 25]) does not fit adequately in any pattern of this survey, but has its *best* fit here. This last match would include and rest upon matches of Slow dark adaptation (good) and Slow speed of judgment (fair) while Low ratio of initial to final performance (M.I. 179, 180)[4] and the measures of Unduly high and Unadaptable (goal discrepancy) aspiration (M.I. 114, 148) level agree poorly, being markedly loaded in his rotation, but almost negligibly in all of these studies.

Possibly the restraint factor changes form when it appears as Dysthymia in neurotics, but we suspect, from our own abnormal group, C4, that the pattern as given here is more truly representative of *all* groups when the factors are thoroughly extracted and rotated. It is interesting to note that this factor strongly supports the theory put forward twenty-five years ago (4) that one of two major determiners of magnitude of PGR deflection is the amount of restraint exercised on orexis, not the conation (or even "emotion") *alone* as commonly supposed (63).

U.I. 18. Hypomanic Overcompensation. Again the match was discovered without effort, a typi-

[4] M.I. will henceforth be employed as abbreviation for Master Index.

TABLE 6

FACTOR 18. HYPOMANIC OVERCOMPENSATION

Master Index No.	Title	Factor in Series: / Sample: / Research:	Factor Identification 5(−) 370 C1	10(−) 370 C2	10 100 C3	2 100a C4	16 500 C5	13 250 C6	13 184c C7
153	High speed of decision on C.M.S. test				69	41	25		10
15	Excessive use of circles on C.M.S. test		11		07	77	31	23	52
39b, 136	High ratio color to form in sorting		32	32	41	30	03	(−06)	
143	High total score on C.M.S.				44	19	18		61
16, 154	Many slanting lines cut on C.M.S.		09		49	01	27	41	49
134	High ratio of self-criticism to criticism of others				15	(−04)	21	47	
193	Ego weakness: shift toward successfuls				08	15	15	38	04
25	Low ratio of emotional to nonemotional words recalled		−21	−03	−33	−20			
266	Low ratio of initial to final performance on C. M. S. or counting				26		45		40
111	Much self-reference in general events				(−06)	40	04	05	02
38	Low ratio of consonant-dissonant opinion recall		−41	−60	−09		−03	−10	
53	High ratio of self-criticism to self-approval						59		
82	High speed in finger maze			58					

cal inter-r (that of the longest common series of variables, see Table 3) being +0.47. However there is to be noted an initially confusing, "cooperativeness" (7) of this factor with factor 16, which shows itself most clearly in the interchangeability of C1: F.5(−) and C2: F.10(−) and C1: F.1(−) and C2: F.3(−), respectively, though the two factor pairs are clearly distinct in other loadings.

There is a tendency for this pattern in the abnormal population to be out of line with the others, e.g., on 38, 134, which is perhaps explicable by its meaning which suggests a compensatory hypomanic pattern of wholehearted action with insecurity and underestimation of the self. Such an overactive pattern might develop new qualities as it approaches abnormal and possibly exhausted extremes, as clinical observation suggests. The variables of ego weakness and conflict over the self distinguish it from the decisive self-assertion of factor 16.

U.I. 19. Critical Exactness. At this point in our series there begin to occur rather more instances of low loadings in some studies for a variable that has a high, salient loading in apparently the same factor in another study; but the sign inversions are still few (7 out of 63) and occur with insignificant values. In one study—C2—there was some rivalry of factors 7 and 8 with factor 5 for the present match, but in addition to exceeding the former in number of congruent variables the latter fits far better here than it would anywhere else. It is interesting to note that when this factor was first found, in C1 (8), the published interpretation predicted it would have relation to questionnaire factor I (25), and this hypothesis was tried and found correct in C5, though not recorded in the present table because we are confined to objective tests.

Our judgment that Thurstone's (61) perceptual factor E belongs here is based on all three common variables fitting, and on the fact that none of his other factors have the least resemblance. Nevertheless, Hidden Pictures would not have been recorded in our series (too low) were it not for its significance in his factor pattern. However, all the signs

TABLE 7
FACTOR 19. CRITICAL EXACTNESS

Master Index No.	Title	Factor Identification: Factor in Series: / Sample: / Research:	7(−) 370 C1	5 370 C2	5 100 C3	12(−) 100a C4	14 500 C5	6 250 6C	7 184c C7	E 190 T2
146	High per cent accuracy in perceptual gestalt completion				55	21	33	48	12	
71	Few errors in maintaining complex reaction time sets			−58			−13	−17	(09)	
51	High index of carefulness			45			37	32		
16	Few slanting lines crossed in C.M.S. Test		−11		(07)	−04	−23	−34	−02	
2	Low motor (and motor-perceptual) rigidity		(01)	(11)	00	−26	−20	−32	−17	
15	Restrained use of circles		−28		−04	(09)	−09	−21	−15	
39	High ratio color to form in sorting		07	28	11	(−16)	31	12	36	
206	Good analysis performance on Gottschaldt figures						25	33		37
41	Good two-hand coordination		34	01	40	10	04		04	59
134	High criticism of others relative to self				−38	−18	(11)	−05		
116	High critical hostility of judgment				57	13	03		11	
80	Little upward drift of PGR resistance relaxing			−38					−13	
170, 198	Hidden pictures and words (mean loading)				06	19	05		11	34
79	Tendency not to have eidetic imagery			−35						
92	Slowness to imitate animal sounds				42					

are consistent. Again we shall argue that its classification elsewhere as a perceptual factor represents (under Thurstone's title "suppressive manipulation of configurations") the outcropping of a broader temperamental quality of exactness, firm control, criticalness, and independence, appreciably associated, incidentally, with masculinity of sex.

U.I. 20. Sociable Willingness. The average matching r was 0.31 and some reduction of what is intrinsically a higher average must be specifically traced to the child population C7 and to the cheating measure (M.I. 16, 20) which was experimentally modified. (It is, in any case, less hardy than most with change of minor conditions, e.g., motivation.) The particular emphases in the C7 study may be connected with the coincidence that this factor generally concerns pervasive attitudes to social values and occurred here in a group which was an orphanage group. Apart from these special deviations the matching here is still substantially at the level found in our first factor.

The meaning is clearly one of dependence upon and willingness to please an emotionally approved authority. Brogden's factor, which he called Honesty, could match either factor 17 or here, but, as far as the limited number of common variables permits judgment, it is better here and the "honesty" is then one of compliance. The factor has some significant connection with femininity, though our earlier speculation (that it might be simply *the* sex factor) is not supported by the magnitude of sex correlation.

U.I. 21. Energetic Decisiveness. The confidence of matching remains very good. The r's between the six successive pairs of factors (C1: F.8 with C2: F.7; C2: F.7 with C3: F.4, etc.) are +0.42, +0.44, +0.35, +0.29, +0.71 and +0.35. The corresponding s values are confidence levels of 0.06, 0.06, 0.03, 0.10, 0.003, and 0.46.

Some resemblance has to be remarked on one set of variables between this pattern and that of factor 20 above (but reflected) as well as to the "ability" factor 6 (in the

TABLE 8

FACTOR 20. SOCIABLE WILLINGNESS

Master Index No.	Title	Factor Identification: Factor in Series: / Sample: / Research:	3 370 C1	1 370 C2	2 100 C3	5(−) 100a C4	6(−) 500 C5	3 250 C6	4(−) 184c C7	2 100 Br
152	Tendency to indicate agreement rather than disagreement				10	68	62	74	67	
135,38	High ratio of recall of consonant to dissonant opinions		03	64	17	39	53	71		
219	Willingness to admit common peccadilloes						−27	−36	−47	
35	Much suggestibility to authority		35	04	29	22	44	36	(−16)	
34	Immaturity of opinion		−09	−04	(+12)	−21	−35		−18	
193,194	Ego weakness, mean loading on shifts to success (193) and from neurotics (194).				09	15	31	25	(−01)	
246	Much authority submission ("authoritarian" attitudes)						−18	−48		
116	Hard-headed cynicism				(−06)	11	11		56	
67	Extremity of viewpoint			01	(−12)	07	(−05)		46	
163,165	Relatively little loss of performance with disapproval (Mean of 110 and 112)				−53	−31	(+07)	00	−16	
46,137	Little impairment of maze performance by shock		−03		−61	(+18)			−02	
36	Inability to state logical assumptions		−44	−51	−03	−02	−03			
21	Absence of questionable preferences in reading		25	57	12	11	15		15	34
16,20	Absence of tendency to cheat. Few slanting C.M.S.		−01		−41	−08	−01	−01	(05)	−60
75	Little metabolic rate change with stimulation			−51						
70	Rapid appearance of negative after image			−43						
294	Good contact with reality						49			

Universal Index [14]). (The first resemblance shows more clearly in factors C1: F.8; C2: F.7; C3: F.4 than in the rest.) The immaturity, suggestibility, etc. set of variables, however, are associated in U.I. 20, with compliance and good behavior, but here with weakness in memory, in computing, in decision, etc. A second direction of "cooperativeness" of the present factor is to be noted here with factor I, Intelligence, on the M.I. variables numbered 167, 199, 244, and 277.

We are inclined to interpret this factor as the cyclothymia pole in the cyclothymia-schizothymia dimension, but while awaiting external validation we have kept more cautiously to the descriptive heading, indicating a certain mental energy and decisiveness as opposed to the schizothyme lack of these qualities. Actually the factor recently found by Eysenck (32), in his search for a cyclic-schizophrenic distinction, matches well here (and nowhere else) on the four common variables that have salient loadings in his study; but his significant but lesser loaded variables—slow speed of reversible perspective and tempo of tapping—have doubtful loadings, and we are inclined to ask whether his values may not be due to the special properties of "criterion rotation."

It seems reasonably clear that the factor studied for thirty years under the title "fluency of association" (or sometimes "speed," "imagination," or "verbal fluency") by Hargreaves (39), Johnson (46) (as a pattern element), Johnson and Reynolds (45), Gibb (36), Carroll (3), Sen (56) (as an influence on Rorschach), Holzinger and Swineford (42), Thurstone (60), Denton and Taylor (30) (as

TABLE 9

FACTOR 21. ENERGETIC DECISIVENESS

Master Index No.	Title	Factor in Series: / Sample: / Research:	Factor Identification 8 370 C1	7 370 C2	4 100 C3	15 100a C4	3 500 C5	1 250 C6	10(−) 184c C7	2† 200a E2	T. XXVII in (7) Composite
34	Maturity of opinion		−25	−42	−56	−66	−08		−43		
35	Little suggestibility to authority		−23	−50	−07	−28	−08	−13	−26		
244,3	Fast speed of judgment (social-emotional)		04	07	(−04)	14	48	70	41		50
167	Good immediate memory				03	47	13	44	06	41	
108	Little assumption of skill in untried performance				−34	−01	−15	−16	−06		
146	Low percent correct in perceptual gestalt completion				−03	−09	−24	−13	−49		
271,1	High fluency (mainly expanding topics)		13	03			32	35	13	54	55
288	Fast speed of evaluating others' performance						51	70			
308	Fast speed of number comparison						48	44			59
95	Much decrease myokinesis under threat				57	20					
199	High computational skill						33	35	09	35	
157	Few jokes considered funny				−38	−08					
152	Tendency to indicate agreement rather than disagreement				15	13	37	24	07		
193	Ego strength: small shift to "successfuls"				(07)	(14)	−06	−09	−54		
33	Myokinesis, large movements*				10					45	
28	High dynamic momentum		(−04)		(−08)	(−15)				16	38
268	Fast natural tempo of tapping						14		(−15)		54
43	Large PGR. reaction to general stimuli		05	01	25	01			(−02)		30
278	Fast reading tempo						59				
309	Fast speed of line judgment						54				
84	Much increase of pulse rate under stress			−53							
74	High basal metabolic rate			34							

* Fairly frequently, in researches outside our own series, some lesser degree of agreement of loadings will be found because tests have not been administered as uniformly as in our series. This is a case in point: Eysenck's test (32) was size of writing; ours was size of lines, freely drawn. Similarly, in Factor T.XV. Study T-R, M.I. 8 in Table 10 a discrepancy in details of presentation exists.

† Due to uncertainties of rotation we have taken the mean of Eysenck's Tables VII and VIII (Factor 2) (p. 382) (32) in computing the factor pattern.

"ideational fluency," not "W"), Line and Kaplan (47), Notcutt (51), McNemar (48), the present writer (5), and others, and represented in the earlier summary (7) with composite loadings as "T.XXVII," has been a perception of this very personality factor. But most of these earlier factors were mutilated, by restriction of tests, to concepts current in ability study and to sets of educational variables. Examination should also be made,

with more common variables, of similarity to the "originality" factor of Guilford and coworkers (38), which combines low gestalt accuracy with fast judgment and excellence in some kinds of fluency. However, this could also match U.I. 19 (reversed) and is considered by Guilford to be additional to ideational and word fluency factors.

Consonant with our interpretation, "fluency" factor measures were shown in 1935 (58) to have substantial relation to manic, cyclic insanity while the "versatility" measures in Denton and Taylor (30) manifested clear relations to the personality questionnaire 'surgency' measures. The reasons for shifting the emphasis to a temperamental interpretation are discussed elsewhere (15).

U.I. 22. Nervous Alert Reactivity. The correlations in this quite broad match are a little lower and more uneven than in former factors. The last two common test variables (excluding sex) that are entered for comparison with the "persistence" factor of Thornton, Ryans, and others are not completely identical with theirs and are made by combining two tests. Thurstone's perceptual factor C agrees on his

TABLE 10

FACTOR 22. NERVOUS, ALERT REACTIVITY

			Factor Identification								
		Factor in series:	4(−)	6	15	17	12	10(−)	9(−)	C	3(−)4(−)*
Master		Sample:	370	370	100	100a	500	250	184c	190	189c
Index No.	Title	Research:	C1	C2	C3	C4	C5	C6	C7	T2	T-R
176	Fast irregularly warned reaction time		−01	−16	−48	−18	−25	−56	−72	−71	
263	Low variance in reaction time							−39	−63		
5	Low ratio of irregular to regular warned reaction time		12	17	(−10)	38		64	31		
94	High frequency of flicker fusion				52	50		11		11	
40	Many reactions to false signals		14				22	02	38		
83	High fidgetometer frequency			43	08	41	14		08		
93	Little effect physical restraint on mental performance				25	39					
159	High accuracy (low scatter) of "time required" estimates				−13		−26	−17			
282	Many objects perceived in unstructured drawing						24	27			
13	Much oscillation of performance		48	03	18	09			07		
197	Few hidden pictures found				−18		(03)		(05)	−16	
8	High speed alternating perspective		33	31	20	(−06)	07		17		31
28, 204	Low dynamic momentum		00		−05	−23			−28		−46
6, 128	High ideomotor and reading speed (or "Fast ideational tempo")		−24	−27	−20				−19		−30
64	Female sex				−25					−22	−50
79	Good eidetic imagery			37							
45	Endurance, time *S* will endure shock and fatigue										−61
325	High level of experience in diverse fields						44				

* T. XV. in (7).

two highest variables well, and tolerably on the two next, but misses a perfect match by "reversible perspective" failing to have any loading on this factor.

A similar pattern has recently been found by Horwitz (43) in so-called rigidity variables which he justifiably matches with the factor of "ideational rigidity" found by Tiner (27) and the present writer. This factor which conceived as "rigidity" is opposite in direction of measurement to the present "over-reactivity" loads, on the two studies mentioned, low flicker fusion, low dynamic momentum (sometimes called "cosatiation"), and low speed of alternating perspective, which might be added to the present pattern. There is disagreement over the sign of hidden pictures, but the present "Objects seen in unstructured drawings" might be more equivalent to Horwitz's scoring of hidden pictures. Or, alternatively, his smaller number of cases, fifty, may account for an errant loading there.

The prevailing feature of this factor is high nervous speed. It is consistent with our interpretation that where we included both unwarned and forewarned reaction time tests the former is more highly loaded. There is thus a general tendency to very alert, sensitive reactivity; to reactions at "half-cock"; and to rapidity of all simple mental processes, along with low endurance or momentum. In a further development we have hypothesized this to be a physiological, perhaps high acetylcholine or hyperthyroid, condition of the neural tissue of the whole nervous system (15).

U.I. 23. Neural Reserves vs. "Neuroticism." This is the last of the factors in which what seems an adequate match over seven or more studies can be made. Its precision is slightly marred by the italicized loading being based on somewhat different scorings or administrations of the common tests indicated. Thus Eysenck scored rigidity on an a priori pattern

TABLE 11

FACTOR 23. NEURAL RESERVES VS. "NEUROTICISM"

		Factor Identification						
	Factor in Series:	10	8	7	7	5(−)	1	1
Master	Sample:	370	370	100	100a	184c	105a	98a
Index No. Title	Research:	C1	C2	C3	C4	C7	E1	E3
2 Low motor rigidity (classical perseveration)		−15	−29	−58	−20	−37	*−42*	
41 Good two-hand coordination		02	20	(−15)	37	36		42
42 Low body sway suggestibility		−20	−13	−26	−12	−10	−64	−12
120 Low ratio of inaccuracy to speed				−20	−29	−36	−40	
151 Slower decision on principles than particulars				31	27			
28,223 High "dynamic momentum" (endurance)		−01		*12*	*36*		*50*	16
43,1 Smaller PGR to mental than physical stimuli (threat)			−07	−31				
101,148 Little excess of aspiration over achievement				−29	−02	−56	−17	
162 High ratio of normal to shock performance				26	46	26		
1 High fluency		00	09			10	40	
167 Good immediate memory				14	(−01)	01	00	33
216 Low suggestibility to shock						−17		−15
51 Low index of carefulness			−67					
75 High metabolic rate change with stimulation			57					
58 Much myokinetic drift			31					
227 Preference for color over form in pictures						36		

Note.—Loadings in italics were not made by variables measured by exactly the same procedure or scoring as the others.

TABLE 12

FACTOR 24. ANXIETY TO ACHIEVE VS. ASSUREDNESS

Master Index No.	Title	Factor in Series: Sample: Research:	Factor Identification 1 500 C5	9(−) 250 C6	12 184c C7	5 100 Br
211	High annoyability: all situations		52	39	21	
219	Willingness to admit common peccadilloes		−42	−19	−43	
152	Tendency to agree		24	63	49	
21	Absence of questionable reading preferences		33		09	36
116	High ratio hard-headed (cynical?) to sentimental attitudes		28		48	
276	Good school grades (intelligence partialled out)				18	40
205	High emotionality of comments		35	13	17	
144	High agreement with platitudes		10		46	
108	Modesty on untried performances		−11	−03	−20	
246	Disagreement with authoritarian attitudes		(05)	(06)	−44	
274	High verbal ability on Stanford Achievement				19	55
330	Good ethical choices in story situation					44
327	Logical inconsistency in responses		51			
294	Low "contact with reality"		−43			

derived from the present writer's finding of low emotional stability with *extremes* of rigidity (5), instead of in the simple, linear fashion of the other rigidity scorings. Incidentally, such special scorings may sometimes adapt to, and give better loadings for, one factor, but make the essential test operation less negotiable and understandable in *all* its factor relations. Our Dynamic Momentum, 28, is the nearest thing to his Endurance, while 223, as used in C1, is identical with it.

This factor is of special interest as being that to which Eysenck's laboratory has devoted its central and constant studies, enriching it with extensive external validations as General Neuroticism (reversed here) (31, 33). Nevertheless, it must be noted that in his first study it was left at an unrotated definition, while in his second and latest it is rotated by a different (but potentially congruent) principle from that of simple structure. In the upshot his first agrees quite well with our five, but his second does not agree so well either with his first or our five as these do among themselves. In particular we cannot find either Fluency or Immediate Memory to be so much involved in this pattern as he finds them, and the same is true to a lesser degree of Speed of Reading and Speed of Judgment.

Our title suggests the hypothesis that the factor is some kind of cortical exhaustion condition produced by neurosis rather than the pattern of poor dynamic integration itself.

U.I. 24. Anxiety to Achieve vs. Assuredness. Despite being found in only three (the latest) C studies the r and s values of matching are good. (Mean $r = +0.47$; above the average for all patterns.) Brogden's "Achievement & Positive Attitudes Factor" also fits pretty well. In study C5 it is shown that this factor correlates 0.6–0.7 with the 16 P.F. "nervous tension" factor Q_4 (25) and with the other anxiety factors O, C− and Q_3−, possibly to an extent required by their intercorrelations. The marked association of this tension and irritability factor with school grades (with intelligence partialled out), verbal ability, and good ethical choices suggests that it is the factor which is responsible for the low correlations recently found between learning (as conditioning) and undifferentiated "anxiety" measures. We shall hypothesize that this is the anxiety-strain factor specifically concerned with degree of long-circuiting of satisfactions, i.e., a measure of the degree of operation of psychodynamic law 10 (11, p. 643).

U.I. 25. Accurate Realism vs. "Psychotic Tendency." We had independently named this factor "Accurate Realism" (or "Calm Decision" in other versions) when we discovered in the matching process that it seems to be the same as Eysenck's "psychoticism" factor reversed. However, the extent of agreement in matching, e.g., the magnitude of r's between patterns, now begins appreciably to decline for the first time in our series. There is some question of whether Eysenck's alternating perspective was scored in the same direction as ours. This might be a borderline match, not rightly included in our first series of 12 most invariant factors were it not that it throws light on and brings support to Eysenck's important contention that a factorizable dimension, definable as psychosis, exists independently of that of neurosis.

TABLE 13

FACTOR 25. ACCURATE REALISM VS. "PSYCHOTIC TENDENCY"

Master Index No.	Title	Factor in Series: Sample: Research:	Factor Identification 16(−) 100 C3	11 300 C5	12′* 184c C7	−1† 105a E2
7	Slow speed of perceptual gestalt completion		−58	−32	−36	
199	High computational performance			35	29	41
8	Slow alternating perspective		−16	−18		−29
146	Low accuracy of gestalt completion		−34	−21		
6	Fast ideomotor (reading) speed		14	03		41
247	Small size of drawing (low expansiveness) copying design		(02)		−15	−61
268	Fast tapping speed			10		21
320	Shorter estimate of time taken working on mirror drawing				−13	−33

* Both a 12 and 12′ (see U.I. 25) have been reported in the C7 study (22) because two clear hyperplanes exist, and it is probable that a thirteenth factor should have been extracted to accommodate this last, 12′ being really a factor 13.

† Owing to uncertainties of rotation, we have taken the mean of Eysenck's tables VII and VIII (Factor 1) (32, p. 382) in computing the factor pattern.

U.I. 26. Cultured, Introspective Self Control. The correlation of those factor patterns with the longest testable common series of variables here, C5 and C6, is +.57; but for the remaining matches the table shows more uneven and low loadings than previously encountered, including a sign inversion of moderate size (on Criticalness of Judgment, though both of the conflicting values are too low to enter here). Consequently, though the pattern is in other respects well enough established to deserve a higher rank, uncertainties of matching put it here.

The pattern indicated by the C studies clearly resembles one found by Brogden (1), Crutcher (29), O'Niel (53), Rethlingshafer (54), Ryans (55), and Thornton (59). These are represented here by the loadings of the most typical—Brogden's. Actually the match is only moderately good, due to the discrepancy on the Dynamic Momentum test, but it is strengthened by the fact that no other factor in our comprehensive series has any resemblance to this important factor in past studies (labeled XVI in the 1946 survey [7]). Some disagreement on M.I. 2 might be due to

TABLE 14

FACTOR 26. CULTURED, INTROSPECTIVE SELF-CONTROL

Master Index No.	Title	Factor in Series: Sample: Research:	Factor Identification 14(−) 100 C3	4 100a C4	8 300 C5	14 230 C6	11 184c C7	3 100c Br
273	High proportion of fluency on self				51	62	10	
271	Low total fluency on topics				−24	−06	−09	
161	High performance under approval relative to disapproval		(−08)	34			31	
28	High dynamic momentum (final)		(−07)				11	69
167	Good immediate memory on attitudes		−04	(−22)	30	13	23	
2	Low motor rigidity		20	−10	−09	(14)	−10	−33
13	Low oscillation		−25				−09	−52
205	Much emotionality of comment				21	03	44	
105	High ratio of threatening objects seen in unstructured drawing		22	(−14)	19	20	14	
39	Preference for form over color in sorting		−03	−23	−01	−21	(07)	
24	Much improvement initial to final period in novel performance		−30	−24		−24	(02)	
266	Low ratio initial to final performance in general material		−35	−24			−06	−53
71	Few erroneous reactions in complex R.T.'s				(23)	−12	−22	
276	High school grade						10	37
36	Ability to state assumptions		32	17	10			
96,188	Little effect of frustrated preference on computational speed		−06	−71				
118	High sense of poetic style				22			
242	More annoyed with things than people						50	
	Zeigarnik ratio							41
	Motor inhibition (Downey)							30

peculiarities of scoring in some of the early studies. It must be borne in mind also that most of these earlier studies were preoccupied with, and were alone heavily represented in, the behavior of dynamic momentum ("cosatiation") as shown by absorption in a task and indifference to interruption, in which area we had only one representative variable, which happens not to have very high loadings anywhere in our studies. The last column match may therefore be considered tentative,

TABLE 15

FACTOR 27. APATHETIC TEMPERAMENT

Master Index No.	Title	Factor Identification		
	Factor in Series: Sample: Research:	17 100 C3	18 100a C4	4 150c Cn
42	High body sway suggestibility	51	66	16
122	High static ataxia	55	35	25
43,1	Low ratio of mental to physical stimulus reaction, P.G.R.	−12	−39	
145	High responsiveness of aspiration level to performance level	51	14	
101	Little excess aspiration level over performance level	−22	−29	
115,193	Marked ego weakness (toward successfuls)	47	(−20)	
194	Marked ego weakness (away from neurotics)	28	12	
13	High oscillation	27	01	22
41,177	Low manual dexterity (two hand coordination and motor speed	−10	(05)	−16
157	Few jokes reacted to by overt smile		−29	
	Low sociometric rating*			−39

* Though this is not S's *behavior per se*, it is included as the highest single loading in Connor's study (28).

though the general sense, as seen in the independently derived labels of the different studies, agrees in depicting a pattern of introspective, steady, serious, self-absorbed, effective thinking. Yet another instance of this factor has recently been provided by Horwitz (43).

U.I. 27. Apathetic Temperament. The agreement here (except for a possibly erroneous, no longer checkable sign on test 193) is very good. The calculation of the marker coincidence index for the C3 and C4 studies shows 1 chance in 500 of so good a match accidentally. The factor of Connor arises from his specific study of short-term and long-term oscillation (28, 49), and though his loadings are not high, it fits appropriately here in pattern and in meaning. Factor 27 must be considered the last and least of the factors for which invariance can be claimed, for it rests on only three independent studies and is the smallest factor in variance in each of these. For these reasons its character cannot so confidently be hypothesized; but it seems to consist of a despondency, apathy, and possible anxiety, associated with high oscillation. The possibility should be investigated that it is Factor C, Ego Strength (11, 25) reversed.

Six further factors can be located in present objective personality measures, but for reasons of space these more tentative factors, with less adequate matching or fewer studies or fainter variance, are reported elsewhere (15). They consist of: U.I. 28, Sociable Emotional Evasiveness, found in five studies; U.I. 29, Sympathetic Mobilization of Energy, found in seven, but with small variance; U.I. 30, Stolidity, found in four; U.I. 31, Wary Realisms, found in three; U.I. 32, Schizothyme Inhibitedness, found in two; and U.I. 33, Dourness, found in two.

DISCUSSION

Although our objective of crystallizing the patterns evidenced in the realm of objective tests has been achieved, to the extent of demonstrating twelve to eighteen tolerably invariant factors, the minds of most readers will doubtless reach ahead to such questions as interpretation and the relation of these factors to those found in the behavior rating and questionnaire media. Space absolutely forbids the amount of discussion needed for the former, but a word may be said on the latter. The pioneer study intercorrelating factors from the three media (24), as well as a second study (24a), revealed a more complex relationship than was anticipated in the theory that the same dimensions of personality would show themselves equally clearly in all three media. Most of the questionnaire factors in the Sixteen Personality Factor Questionnaire align themselves with factors in the rating of behavior *in situ*; but as between these factors and the laboratory test factors a rather confused relationship obtains. It is possible that the conceptual difficulty will resolve itself by some factors in the objective test proving to be second-order factors among those in the rating-questionnaire realm. For it has already been shown that the four distinct questionnaire anxiety factors correlate significantly with, and only with, the objective test anxiety factor U.I. 24; while three questionnaire "introversion-like" factors similarly center upon U.I. 32, Schizothyme Inhibitedness (19). Advance toward clarifying this complex issue can now be made only by that more accurate determination of the factors and their obliquities, on large samples, which will permit effective second-order factorization.

Parenthetically, at the conclusion of this survey, one may call attention to the fact that even for psychologists who do not follow the

logic and implications of factor analysis, these studies contain some 35,000 correlations among variables operationally representing almost every concept used in personality research. The practice whereby psychologists create armchair hypotheses, from observations and notions no more exact than those which they share with the man in the street, is fortunately passing out of fashion, wearied by its own sterility. The new hypotheses that will bring the advances of the future must spring from minds thoroughly immersed in a wide field of exact, fresh, personality-relationship data. Such a domain exists in the records of correlation matrices here surveyed, and a careful perusal thereof would frequently save unnecessary experimentation and make possible, initially, more informed and incisive hypotheses.

Attention evidently needs to be called to this possibility, not only for the sake of generating better hypotheses, but also in terms even of completeness of supposed exhaustive experimental "data" surveys. For example, several recent "surveys" covering such topics as the associations of the GSR, the nature of rigidity, tempo, ego strength, responsiveness to humor, fluctuation of attitudes, test response sets, suggestibility, flicker fusion, etc. have completely overlooked the extensive correlational evidence available in factor analytic publications, simply because references to specific tests do not customarily appear in titles of factor analytic articles. Yet the latter studies generally include richer experimental data, and frequently better population samples, than those occurring in the whole bibliography of the piecemeal univariate studies listed.

However, the main scientific advantage of this multivariate experimental approach is that the data are integrated and structured—if the evidence of this article holds—more than is possible by piling univariate upon univariate study, in no matter how assiduous a fashion. Its disadvantage is that it assumes linear relations, but not one univariate study in a hundred deals with any more complex relation! For those who are interested in such concepts as anxiety, rigidity, ego strength, speed of conditioning, oscillation, aspiration level, frustration tolerance, authoritarian attitudes, etc. the present patterns are a challenge to climb off subjective concepts on to firmly based evidence of functional unities—which may sometimes require a salutary revision of the subjective theoretical obsessions. Operational representatives of all the concepts just listed, and more besides, have been included in the 330 variables of the Master Index, so whatever unitary patterns and concepts actually exist among them have had a chance to manifest themselves.

Methodologically, since scientific concepts (as distinct from those of philosophy) have to grow out of empirically-observed, lawfully-recurring patterns such as these, the derivation of personality concepts by multivariate, factor-analytic approaches is actually more economical too. For the alternative is to follow the long trial-and-error path of discarding, through a painful history of unsuccesaful univariate experiments, a succession of concepts, each of uncertain or subjective origin. For example, the confident clinical conceptualization of rigidity as a unitary quality of the individual to which the present writer once subscribed (5) must be abandoned, except as a very narrow factor in a set of very similar and largely motor tests. A very large number of univariate, controlled experiments failed to reveal that it needed to be abandoned. The present survey shows four, possibly five (U.I. 20, 21, 22, 29, and 32) personality factors (besides low intelligence) contributing to rigid behavior in the clinical sense, and it is now easy to see the *kind*, i.e., the psychological quality, of contribution to such vaguely conceived rigid behavior that each makes. Similarly, though U.I. 24 is the clearest form of anxiety, one can see other and independent dimensions, e.g., U.I. 23(−), which deserve consideration and clearer formulation in the general study of anxiety. In short, personality factors are not complete interpretations, but they *are* reality-guaranteed concepts around which experiments on personality and deeper interpretation can be more profitably organized.

SUMMARY

1. From some 150 factors in objective personality tests, the loading patterns of which have been published (104 of our own and 46 of various other researchers in this field), we have found, by cross matching in all possible ways, what appear to be 18 potentially invariant patterns. They may be divided, ad-

mittedly at a somewhat arbitrary point, into 12 of a satisfactory degree of invariance and universality, and 6 of lesser statistical significance. The former, reported here, cover 90 phenomenally observed (7) factors (74 in our seven core studies, 16 from studies by others).

2. A development has been carried out in techniques of matching, by demonstrating the r values typically found for a well-known, "established" factor (intelligence) and by the invention of s, the *salient variable similarity index*, based on quite different assumptions from r, but agreeing substantially in verdict. Both indices have been used here, but the art of matching only begins with indices and requires consideration also of (*a*) adequacy of rotations; (*b*) nature of the sample of population, including effects on test meaning, reliability, means, etc.; (*c*) mode of factor computation; (*d*) number and nature of common markers; (*e*) test administration conditions; and (*f*) possible alternative matches. Over the 18 matched factor matrices (12 here; 6 elsewhere [15]), the r's range from 0. to 0.95, mostly 0.25 to 0.75, centering on 0.4, and the s's range from the 45 per cent to well beyond the 1 per cent level (mostly in the 15 per cent to .01 per cent range). The values are about the same for the new personality factors as for the long-known general ability factor and are quite different from the distribution for factors paired randomly.[8]

3. Our immediate purpose has been to reveal the *patterns* for operational reproduction, not to attempt interpretations, which would require more space for discussion and allusion. These patterns are for the young adult population. The abnormal and child populations appear to be essentially the same, but have some interesting differences which at this time must be considered indicated rather than proved. It is noteworthy that the apparent failures of invariance in the present data are systematically greater for the child and abnormal populations. The serious discrepancies, namely the sign inversions of loadings, constitute the following percentages of the 604 matching loadings entered in the seven C studies: C1: 11 per cent, C2: 8 per cent, C3: 12 per cent, C4: 12 per cent, C5: 8 per cent, C6: 8 per cent, C7: 14 per cent. The child and abnormal samples (plus C3 which has a small population) thus again have the greatest tendency to diverge. Consequently, these are probably not failures of invariance, i.e., blurring of the pattern by error, but indicate changes in the real pattern with changes in population species.

5. Discouraging experiences in earlier attempts to match (7, 35) with known certainty among published "isolated" factor analytic studies clearly show that the establishment of invariant, stable personality factors is not to be achieved by improved matching techniques and good marker tests alone, but requires attention to the social organization of research. In particular, we need more long-term projects, with gradual, strategically-planned expansions to new population species and systematic cross checks of structure on different samples, by sufficiency of common markers. When isolated or exploratory individual studies are undertaken there needs to be far more attention to carrying a sufficiency of exact markers from known factors, and to carrying simple structure determination blindly to *the highest possible degree of finish*. The external validation of factors in school, clinical, industrial, and social psychology belongs to a stage after this definition of factors per se.

6. The patterns presented here may be regarded in themselves as hypotheses to be tested—hypotheses concerning functional unities or lawful recurrences of form. In investigating these hypotheses with the aim of proceeding to a precise check on each pattern and to further interpretation, through discovering more highly loaded variables than any yet known, psychologists may now concentrate on single factors and therefore on lengthened tests requiring more time-consuming experimental procedures. From the general nature of any factor, as it now stands partially revealed, it should be possible to modify most tests in a direction that will raise their correlations with the factor.

The derivation of the single "ideal" pattern (in existing variables) from the composite

[8] Since no random r's have yet been set out in any table we may list a few taken haphazardly here. C5's with C6's: $r_{1,3} = -.02$; $r_{14,16} = -.12$; $r_{10,11} = .14$; $r_{2,4} = .11$; $r_{3,8} = 12$; C3's with C4's: $r_{9,5} = .11$; $r_{8,5} = .05$; $r_{10,12} = -.02$; $r_{9,17} = .11$; C4's with C5's: $r_{17,1} = .01$; $r_{9,8} = .09$; $r_{13,16} = .16$. From C6's with C7's: $r_{10,15} = .04$; $r_{18,13} = -.16$. It will be seen that they fall roughly in the range one would theoretically expect for essentially zero r's on populations averaging about 45, whereas the r's mentioned in the matching are almost entirely above this range.

tables here presented requires averaging of the given loadings, with some allowance for regression where there are only single loadings and some allowance for the differing actual and possible reliabilities of tests.

7. With the definition of these factors and the development of more saturated measures for them, the way should at length be open for testing hypotheses about their nature, by controlled experiment and by correlation with everyday criteria.[6] Already it is evident that few of these factors line up directly with rating and questionnaire factors (though these tend to be mutually consistent) and that relations next need to be sought in terms of first-order–second-order factor matchings.

REFERENCES

1. BROGDEN, H. E. A factor analysis of forty character tests. *Psychol. Monogr.*, 1940, **52**, No. 3 (Whole No. 234), 39–55.
2. BURT, C. L. *The factors of the mind.* London: Univer. of London Press, 1940.
3. CARROLL, J. B. A factor analysis of verbal abilities. *Psychometrika*, 1941, **6**, 279–308.
4. CATTELL, R. B. Experiments on the psychical correlate of the psychogalvanic reflex. *Brit. J. Psychol.*, 1928, **19**, 34–43.
5. CATTELL, R. B. Temperament tests: II. Tests. *Brit. J. Psychol.*, 1933, **24**, 20–49.
6. CATTELL, R. B. An objective test of character temperament. II. *J. soc. Psychol.*, 1944, **19**, 99–114.
7. CATTELL, R. B. *Description and measurement of personality.* New York: World Book, 1946.
8. CATTELL, R. B. Primary personality factors in the realm of objective tests. *J. Pers.*, 1948, **16**, 459–487.
9. CATTELL, R. B. A note on factor invariance and the identification of factors. *Brit. J. Psychol. Statist. Sec.*, 1949 **1**, 134–139.
10. CATTELL, R. B. r_p and other coefficients of pattern similarity. *Psychometrika*, 1949, **14**, 279–298.
11. CATTELL, R. B. *Personality: a systematic, theoretical and factual study.* New York: McGraw-Hill, 1950.
12. CATTELL, R. B. A factorization of tests of personality source traits. *Brit. J. Psychol. Statist. Sec.*, 1951, **4**, 165–178.
13. CATTELL, R. B. *Factor analysis: a manual for the psychologist and social scientist.* New York: Harper, 1952.
14. CATTELL, R. B. A universal index for psychological factors. Urbana, Ill.: Lab. of Pers. Assess. & Group Behav., Univer. of Illinois, 1953. (Adv. Publ. No. 3.)
15. CATTELL, R. B. *The theory and practice of personality and motivation measurement*, in preparation.
16. CATTELL, R. B. The O-A Personality Test Battery. Champaign, Ill.: Inst. Pers. & Abil. Test., 1954.
16a. CATTELL, R. B. The chief invariant psychological and psycho-physiological functional unities found by P-technique. *J. clin. Psychol.*, in press.
17. CATTELL, R. B., & BAGGALEY, A. R. The salient variable similarity index-s-for factor matching. Urbana, Ill.: Lab. of Pers. Assess. & Group Behav., Univer. of Illinois, 1954. (Adv. Publ. No. 4.)
18. CATTELL, R. B., CATTELL, A. K. S., & RHYMER, R. M. P-technique demonstrated in determining psycho-physiological source traits in a normal individual. *Psychometrika*, 1947, **12**, 267–288.
19. CATTELL, R. B., CHECOV, L. L., COGAN, E. A., MILLER, A. W., & SCHIFF, H. Personality structure in objective tests: a preliminary to the development of psychiatric screening. *Air Univer. Monogr. Rep.*, 1955, No. 9. (Project No. 21-0202-0007.)
20. CATTELL, R. B., DUBIN, S. S., & SAUNDERS, D. R. Personality structure in psychotics by factorization of objective clinical tests. *J. ment. Sci.*, 1954, **100**, 153–176.
21. CATTELL, R. B., DUBIN, S. S., & SAUNDERS, D. R. Verification of hypothesized factors in 115 objective personality test designs. *Psychometrika*, 1954, **19**, 209–230.
22. CATTELL, R. B., & GRUEN, W. Measurement of the primary personality factors in children by means of objective tests. *J. Pers.*, in press.
23. CATTELL, R. B., & MILLER, A. A confirmation of the ergic and self-sentiment patterns among dynamic traits (attitude variables) by R-technique. *Brit. J. Psychol.*, 1952, **4**, 280–294.
24. CATTELL, R. B., & SAUNDERS, D. R. Inter-relation and matching of personality factors from behavior rating, questionnaire, and objective test data. *J. soc. Psychol.*, 1950, **31**, 243–260.
24a. CATTELL, R. B., & SAUNDERS, D. R. Beiträge zur Faktoren-Analyse der Persönlichkeit, *Z. f. exper. u. angew. Psychol.*, 1955, **2**, 325–357.
25. CATTELL, R. B., SAUNDERS, D. R., & STICE, G. F. The Sixteen Personality Factor Questionnaire. Champaign, Ill.: Inst. Pers. & Abil. Test., 1950.
26. CATTELL, R. B., & SCHIFF, H. The prediction of pilot adjustment from personality factor measures. *Air Univer. Monogr. Rep.*, 1955, No. 10. (Project No. 21-0202-0007.)
27. CATTELL, R. B., & TINER, L. G.: The varieties of structural rigidity. *J. Pers.*, 1949, **17**, 321–341.
28. CONNOR, D. V. The relationship between temperamental traits and intelligence test performance.

[6] The present writer has arranged to make available, under the title of The Objective-Analytic Personality Test Battery ("O-A Battery") (16), the tests and instructions for what seem the best combination or pattern of tests for measuring (*a*) the single factors, (*b*) all twelve factors of the present article together, (*c*) all eighteen, i.e., U.I. 16 through U.I. 33 together. In the two last, tests are chosen to avoid overlap among factors and give minimal factor estimate intercorrelations. This O-A battery is primarily for research on factor structure and only secondarily for the factor measurement of individuals. It is obtainable from I. P. A. T., 1602 Coronado Drive, Champaign, Illinois.

Paper read at Brit. Psychol. Soc., Oxford, March, 1952.

29. CRUTCHER, R. An experimental study of persistence. *J. appl. Psychol.* 1934, **18**, 409–417.
30. DENTON, J. C., & TAYLOR, C. W. A factor analysis of mental abilities and personality traits. (In press, cited by French [**35**]).
31. EYSENCK, H. J. *Dimensions of personality.* London: Kegan Paul, French Trubner, 1947.
32. EYSENCK, H. J. Schizothymia-cyclothymia as a dimension of personality. II. *J. Pers.*, 1952, **20**, 345–384.
33. EYSENCK, H. J. *Scientific study of personality.* London: Kegan Paul, French Trubner, 1952.
34. FRENCH, J. W. Selected tests for reference factors. Princeton, N. J. Educ. Test Serv., 1953.
35. FRENCH, J. W. The description of personality measurements in terms of rotated factors. Princeton, N. J. Educ. Test Serv., 1953.
36. GIBB, C. A. Personality traits by factorial analysis. III. *Aust. J. Psychol. & Phil.*, 1942, **22**, 1–27.
37. GUILFORD, J. P. The discovery of aptitude and achievement variables. *Science*, 1947, **106**, 279–282.
38. GUILFORD, J. P., WILSON, R. C., & CHRISTENSEN, P. R. A factor-analytic study of creative thinking. II. Administration of tests and analysis of results. *Univer. Sth. Calif. Rept. Psychol. Lab.*, 1952, **8**, 3–23.
39. HARGREAVES, H. L. The faculty of imagination. *Brit. J. Psychol Monogr. Suppl.*, 1927, **3**, No. 10.
40. HARTSHORNE, H., MAY, M. A., & SHUTTLEWORTH, F. K. *Studies in the organization of character.* New York: Macmillan, 1930.
41. HOLZINGER, K. J., & HARMAN, H. N. *Factor analysis: a synthesis of factorial methods.* Chicago: Univer. of Chicago Press, 1941.
42. HOLZINGER, K. J., & SWINEFORD, F. A study in factor analysis: the stability of a bi-factor solution. *Suppl. educ. Monogr.* No. 48, 1939.
43. HORWITZ, L. Rigidity factors in normals and psychiatric patients. In press.
44. HUNT, J. MCV., & KOGAN, L. S. The need for psychological theory in evaluating the results of psychotherapy. *Psychol. Serv. Cent. J.*, 1950, **2**, 77–82.
45. JOHNSON, D. M., & REYNOLDS, F. A factor analysis of verbal ability. *Psychol. Rec.*, 1941, **4**, 183–195.
46. JOHNSON, W. B. Euphoric and depressed moods in normal subjects, I, II. *Charact. & Pers.*, 1937–38, **6**, 79–98, 188–202.
47. LINE, W., & KAPLAN, E. The existence, measurement and significance of a speed factor in the abilities of public school children. *J. exp. Educ.*, 1938, **1**, 1–8.
48. MCNEMAR, Q. Practice and "general" motor ability. *J. gen. Psychol.*, 1930, **3**, 57–97.
49. MADIGAN, M. E. A study of oscillation as a unitary trait. *J. exp. Educ.*, 1938, **6**, 322–339.
50. MIRA, E. Myokinetic diagnosis: a new technique for exploring the conative trends of personality. *Proc. roy. Soc. Med.*, 1940, **33**, 173–194.
51. NOTCUTT, B. Perseveration and fluency. *Brit. J. Psychol.*, 1943, **33**, 200–209.
52. OATES, D. W. Group factors in temperament qualities. *Brit. J. Psychol.*, 1929, **20**, 118–136.
53. O'NIEL, W. M. See tables from this investigator quoted in Chapter 11 by Cattell (**7**).
54. RETHLINGSHAFER, D. The relation of tests of persistence to other measures of continuance of action. *J. abnorm. soc. Psychol.*, 1942, **37**, 71–82.
55. RYANS, D. G. The meaning of persistence. *J. gen. Psychol.*, 1938, **19**, 79–96.
56. SEN, A. A statistical study of the Rorschach test. *Brit. J. Psychol. Statist. Sec.*, 1950, **3**, 21–39.
57. SPEARMAN, C. *The abilities of man.* London: Macmillan, 1927.
58. STUDMAN, L. G. Studies in experimental psychiatry, 5: "w" and "f" factors in relation to traits of personality. *J. ment. Sci.*, 1935, **81**, 107–137.
59. THORNTON, G. R. A factor analysis of tests designed to measure persistence. *Psychol. Monogr.*, 1939, **51**, No. 3 (Whole No. 229), 1–42.
60. THURSTONE, L. L. *Primary mental abilities.* Chicago: Univer. of Chicago Press, 1933.
61. THURSTONE, L. L. *A factorial study of perception.* Chicago: Univer. of Chicago Press, 1944.
62. THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univer. of Chicago Press, 1949.
63. WOODWORTH, W. S. *Experimental psychology.* New York: Holt, 1938.

Received September 22, 1953.

THE LEVEL OF ASPIRATION OF SCHIZOPHRENIC AND NORMAL SUBJECTS[1]

KENNETH C. JOST

Veterans Administration Hospital, Danville, Illinois

THE term *level of aspiration*, introduced by Tamara Dembo (2), names a method of investigating goal setting and goal striving which has been widely used in psychology. Few investigations, to our knowledge, have employed psychotic patients as subjects. Hausmann (7), Miller (9), and Winder (12) worked with schizophrenic patients, and Escalona (3) worked with manic-depressive patients. Current texts which mention the aspiration behavior of schizophrenic patients routinely repeat Hausmann's report that these patients, other than those of the paranoid subtype, indiscriminately raise their bids after failure and lower them after success, while those of the paranoid type cling to high estimates regardless of their achievement scores and in spite of the fact that they are doubly penalized for falling below their estimated scores. Inadequacies in this study of Hausmann have led the writer to re-examine the behavior of schizophrenics in a typical level-of-aspiration test.

Inasmuch as schizophrenic patients in general show deficiencies in "reality testing" and in their adaptation to the demands made upon them by their life situations, it is to be expected that they will manifest aspiration behavior which deviates from that of so-called normal individuals. This study asks if there are significant differences between schizophrenic patients and normal subjects with respect to the following selected level of aspiration variables: (*a*) success score (number of times *S* succeeded in reaching or exceeding his bid), (*b*) reaction time (number of seconds required for subject to state his initial bid), (*c*) goal discrepancy score (the mean of the differences, i.e., each bid minus the preceding performance score), (*d*) typical response score (the number of shifts of the level of aspiration upward after success plus the number of shifts downward after failure), and (*e*) forgetting score (the number of times *S* failed to state a bid before beginning a new trial).

This study also asks if the findings reported by Hausmann (7) are reproducible. That is to say, do schizophrenic patients, other than those of paranoid subtype, indiscriminately raise their bids after failure and lower them after success, while paranoids cling to high estimates regardless of their objective scores and in spite of the fact that they are doubly penalized for falling below their estimated scores?

METHOD

In general the method has been to have a group of schizophrenic patients and a group of presumably normal individuals perform on a level-of-aspiration test, and to compare the various scores of the two groups.

Subjects. The subjects were 29 white, male, schizophrenic patients and 29 white, male, hospital employees, for the most part, attendants. The patients were selected by the following criteria: (*a*) chronological age not over 45 years, (*b*) ward cooperation (as judged by the patient's physician and nurse) to be sufficiently high to enable him to perform in a testing situation, (*c*) absence of significant physical disorders, and (*d*) patient to receive a course of electroshock treatments beginning on the day following the first administration of the level of aspiration test[2]. The hospital employees were reasonably well matched with this group of schizophrenics with regard to chronological age, educational level, and Barr occupational ratings. None of the differences between group means was significant ($p > .10$ by a t test).

Task and apparatus. Guided primarily by the criteria of novelty, intrinsic motivation, variable performance, inability to gauge performance, and general adaptability which had been agreed upon by Eysenck (4) and Rotter (10), the Rotter Aspiration Board was selected as the task. This Board, described by Rotter in detail (10, p. 413) consists of a board 38 inches long with a square groove down the center. The task consists of hitting a steel ball along the groove with a stick resembling a miniature billiard cue. Regularly spaced depressions associated with numbered units serve to slow down the speed of the ball and to provide a resting place for it when it comes to a stop. The score for a given shot depends upon how close to a central unit the ball comes to rest. This central unit, painted white with a black number on it, counts as ten points. The

[1] From Veterans Administration Hospital, Danville, Illinois.

[2] The purpose of the second administration of the test to these subjects was to study the effect of electroconvulsive treatment upon the level of aspiration, results of which are in process of preparation for publication.

depressions on either side count nine points, the next ones on either side count eight points, etc.

Two alterations were made in this apparatus for the purposes of this study. Following the suggestion by Klugman (8, p. 102), the 16 depressions preceding the numbered units were omitted, thereby assuring *S* of at least one point for each shot. This was done to enhance the interest of the test for schizophrenic patients. Second, hard maple instead of soft pine was used in construction of the board. This made less likely the gradual elongation of the depressions with use, thus insuring constancy with respect to difficulty in scoring.

Procedure. The *E* first introduced himself to *S* and explained (in the case of the schizophrenic patient) that he was to see him for a few minutes at the request of the patient's physician. The *S* was then shown the Rotter Board. As soon as practicable, the directions described by Klugman (8, p. 103), slightly modified in order to simplify the terminology for use with psychotic subjects, were read aloud to *S* and he was instructed to begin his four sets of practice trials.

Following each of the four practice trials, *E* computed the trial's sum, which score was reported verbally to *S*. At the conclusion of the practice period, *S* was told that this part of the procedure had been completed. The *E* then read aloud to *S* the instructions for the experimental period. These instructions emphasize the need to set as realistic goals as possible, encouraging *S* to state what he actually expects to earn on each set of trials, not the most he hopes to achieve and not the least he expects to earn. This phase of the experiment terminated with *E*'s question, "And now, exactly how many points will you actually score on the first set of five shots?"

At the end of this question *E* started the stop watch, stopping it when *S* responded with an estimate or "bid." This interval provided *E* with the reaction time. Each verbally expressed estimate or bid was then followed for twenty trials by a performance set consisting of five hits of the steel ball with the cue. At the twenty-first and last estimate *S* was told to stop. Thus no performance followed the statement of the twenty-first bid.

Following each set of five hits *E* totaled the actual performance scores (from the record of each hit per trial of five) and said, "Well, you made _____ that time. You said you would make _____ so you will get credit for _____." The *E* then wrote the credit score (defined as the aspiration estimate minus penalty —after the scoring method early suggested by Hausmann and later adopted by Rotter) in large numbers on a miniature blackboard where it could be easily seen by *S*. The purpose of this procedure was to focus, it was hoped, attention upon the credit score and thus to try to make certain that *S* knew his score prior to stating his next bid for the succeeding set of five hits.

Each *S* was tested individually and in private. He was also instructed not to discuss the test with others. Throughout the practice and experimental segments of this study *E* recorded the results for each *S* on a ruled record sheet on which the ordinate represented the scores earned on each hit in sets of five and the abscissa the number in the series of four practice trials or twenty experimental trials.

RESULTS

Success score. The mean success score for the normal group ($M = 11.03$, $SD = 2.46$) was not significantly greater ($p > .10$) than the success score mean of the schizophrenic group ($M = 10.89$, $SD = 4.08$). The two groups did differ, however, with respect to variances, the variability of the schizophrenic group, as measured by the F test, being significantly greater ($p < .02$) than that of the normal group.

Reaction time. The two groups differed with respect to both means and variances. The mean of the schizophrenic group ($M = 26.68''$) was significantly higher ($t = 2.70$, $p < .01$) than the mean of the normal group ($M = 5.17''$); the variability of the schizophrenic group was also significantly greater ($p < .02$) than that of the normal group.

Goal discrepancy score. Analysis of the data for this measure was directed toward answering two questions: (*a*) do the two groups differ with respect to the frequency of negative and/or positive, individual, mean goal discrepancy scores and (*b*) do the two groups differ with regard to the frequency of *high* (negative and positive) individual, mean goal discrepancy scores? In the schizophrenic group 52 per cent of the scores were negative and 48 per cent were positive. For the normal group these proportions were 55 per cent and 45 per cent, respectively. Inspection of these proportions reveals that the two groups did not distinguish themselves with regard to the frequency of negative and/or positive, individual, goal discrepancy means. Our second question, however, must be answered in the affirmative. Twenty-four (83 per cent) of the schizophrenics and fourteen (50 per cent) of the normal subjects obtained individual mean goal discrepancy scores which were greater than 1.06, this cutting point being the median absolute size of the 29 individual means in the normal group. The difference between the frequencies yielded a chi square, corrected for small frequencies, of 6.18, which has a probability of less than .02.

It is of interest to compare the range of individual mean goal discrepancy scores for the two groups of subjects. For the schizophrenic group the range is 19.15 (−12.20 to +6.95), while the range for the normal group is only 7.35 (−2.65 to +4.70). Thirty-four per cent

of the schizophrenic patients obtained individual mean goal discrepancy scores which fell outside the limits of the normal group. Of these extremely deviant scores, six fell at the negative end and four at the positive end of the scale.

Typical response score. The mean typical response score of the schizophrenic group, (M = 8.97) (SD = 4.03), was found to be significantly lower ($p < .02$) than the mean of the normal group, (M = 11.68) (SD = 3.28). When the F test for homogeneity of variances was applied to the data, however, no significant difference was found, p being greater than .10.

Analysis of all of the possible level of aspiration movements (580 for each group of subjects) revealed the following proportions. In the schizophrenic group 44 per cent of the responses were *typical* (raisings after success plus lowerings after failure), 12 per cent were *atypical* (raisings after failure plus lowerings after success), and the remainder (44 per cent) were repetitions of the preceding level of aspiration. For the normal group these proportions were 59 per cent, 7 per cent, and 34 per cent respectively. It will be observed that repetition of the preceding level of aspiration occurred 10 per cent more frequently and atypical responses appeared 5 per cent more frequently among the schizophrenic patients than among the normal subjects. Subjecting these observations to statistical analysis, it was found that 70 per cent of the schizophrenics and only 43 per cent of the normal subjects stated the same bid three or more times consecutively. The difference between the frequencies as measured by the chi-square test is significant at the .02 level (χ^2 = 5.51). When the cutting point was raised to ten or more repetitions of the same level of aspiration, 27 per cent of the schizophrenics and none of the normal subjects fell into this category. It is of interest to note that 4 of the 29 schizophrenics repeated the same level of aspiration throughout the entire task.

In view of Rotter's assertion (11, p. 163) that when there are three or more atypical shifts out of the twenty possible LA movements, the chances of emotional instability are very high, it was expected that the frequency of occurrence of such shifts would be significantly greater among the schizophrenic patients than among the normal subjects. This expectation was borne out by the finding that 52 per cent of the schizophrenics and only 24 per cent of the normals responded with three or more atypical shifts during the test situation. The difference between the frequencies (χ^2 = 4.69) is significant at the .05 level. Thus it can be seen that the lower frequency of typical responses among the schizophrenic patients than among the normal subjects is the result of the schizophrenic's greater tendency both (*a*) to repeat the same level of aspiration consecutively and (*b*) to employ atypical shifts.

Forgetting score. Forgetting occurred among 24 (83 per cent) of the schizophrenics and only 12 (41 per cent) of the normal subjects. The chi-square value (χ^2 = 8.86) for the difference between frequencies is significant at the .01 level.

In order to determine if there was any difference between the frequency of forgetting after success and the frequency of forgetting after failure, two ratios were computed for each of the 29 schizophrenic patients: (*a*) the number of forgets after failure to the number of failures and (*b*) the number of forgets after success to the number of successes on the task. No analysis of these forgettings was made for the normal subjects inasmuch as the frequency of occurrence of such responses was negligible for this group.

If the frequency of forgetting is the same after success as after failure, we should expect about half of the differences between these two ratios to be positive and half negative. Employing the "sign test," a nonparametric test of significance described by Festinger and Katz (5, p. 547), it was found that 18 of the 29 differences were positive (ratio number 1 was greater than ratio number 2), 6 were negative, and 5 were zero. The difference in the distribution of positive and negative signs was found to be significant at the .05 level when the formula $[(N - 1)/2] - (0.98)\sqrt{N + 1}$, where N is the number of pairs having a deviation greater than zero, was applied to the data. From these results, the conclusion is drawn that among schizophrenic patients forgetting occurs proportionately more frequently after failure than after success.

Finally an analysis of the reactions to suc-

cess and failure of two schizophrenic subgroups was initiated to compare the findings with those reported by Hausmann (7) in 1933. It is recalled that this investigator, in a study of the level of aspiration with dart-throwing as a task, reported that schizophrenics, other than paranoid in subtype, *indiscriminately* raised their bids after failure and lowered them after success. Hausmann also contends that paranoids *cling* to high bids regardless of objective scores and in spite of the fact that they are penalized for falling below their predicted scores.

If Hausmann's results are reproducible, we should expect to find in the present study of 29 schizophrenics (14 nonparanoid and 15 paranoids): (*a*) a greater proportion of cases in the nonparanoid than in the paranoid subgroup employing atypical shifts and (*b*) a higher aspiration score mean (mean of the first 20 bids) for the paranoid schizophrenics than for the nonparanoid schizophrenics. The results of the present study fail to bear out either of these expectations. In the first place a greater proportion of paranoids (67 per cent) than nonparanoids (50 per cent) gave one or more atypical responses during the twenty trials. The chi-square value ($\chi^2 = .82$) for the difference between frequencies, is not statistically significant, however, p being greater than .30. Second, the aspiration score mean of the paranoid subgroup, ($M = 27.92$) ($SD = 5.41$), is not significantly greater ($p > .10$) than the mean of the nonparanoid group, ($M = 27.25$) ($SD = 5.82$). Thus the results of the present investigation fail to support Hausmann's assertions.

DISCUSSION

As is generally known, there are numerous writers in the broad field of psychopathology who in theory and practice consider loss of self-esteem a most important agent in the development of schizophrenia. Anton Boisen (1, p. 27) holds that the condition represents a reaction to an unfulfilled achievement drive, a condition shown by those who judge themselves as "personal failures." "There is no objective measure of such maladjustments. The individual himself is the judge." This statement, phrased differently, is encountered in the clinical literature with startling regularity. Those who accept this loss of self-esteem as the core of the psychodynamics that gradually or rapidly leads into the eventual withdrawal from a feared environment go on to emphasize that the patient stops maintaining his relationship (in degree) with those about him the better to protect himself from further threats to his self-esteem. Because loss in self-esteem is so painful (Boisen, 1) and because further losses through the medium of additional failures are psychologically catastrophic (after Kurt Goldstein, 6), the schizophrenic patient is more sensitive than the normal adult to situations involving threats of failure. In the light of these assertions, the significant differences which have been found between the two groups in this study are hypothetically held to be the result of a stronger need on the part of the schizophrenic patients than on the part of the normal subjects to maintain and/or bolster self-esteem.

The stronger need to maintain or prevent loss of self-esteem may well explain the higher-than-normal reaction time mean of the schizophrenic patients. It is posited that these patients were more indecisive than the normal subjects about stating the initial aspiration estimate because of their greater reluctance to commit themselves in a novel situation, this reluctance being interpreted, in turn, as an attempt to maintain self-esteem through defense of the ego against failure.

The greater frequency of extremely high (negative and positive) goal discrepancy scores among the schizophrenic patients than among the normal subjects is also attributed to the greater need on the part of the former subjects to maintain and/or bolster self-esteem. In the case of the high, negative goal-discrepancy score the subject attempts to reduce the probability of failure and/or increase the probability of success by setting low explicit goals. The high, positive goal-discrepancy score, on the other hand, is interpreted as a compensatory attempt on the part of the subject to bolster his self-esteem.

The stronger need to maintain self-esteem may also explain the lower-than-normal typical response score mean obtained by the schizophrenic patients. This finding has been attributed to (*a*) the schizophrenic's greater tendency to repeat, consecutively, a previous level of aspiration and (*b*) his greater tendency

to give atypical responses. Both restatement of the previous level of aspiration and raising the level of aspiration after unsuccessful trials represent denials of or reluctance to accept past failure, while both repetition of the previous level of aspiration and lowering the level of aspiration following successful trials represent attempts to avoid future failure. Both of these patterns, it is held, are components of the need to maintain self-esteem.

The greater incidence among the schizophrenic patients than among the normal subjects of forgetting to make verbal statements of goals before beginning new trials is attributed to the greater need on the part of the patients to "leave the field" in the attempt to maintain self-esteem. The schizophrenic patient, more so than the normal subject and because of his past failure experiences, enters the test-situation anticipating further failure. Because of this set, he attempts at much more frequent intervals than the normal subject to withdraw from the threatening situation by not making verbal statements of his goals. In the case of the schizophrenic patient, this need to withdraw from the threatening situation apparently manifests itself more strongly after unsuccessful trials than after successful ones.

No attempt is made here to interpret the significant differences which have been found between the variances of the two groups for the measures *success score* and *reaction time*. Many investigators have found similar differences in interpersonal variability between psychotic and normal groups for a variety of measures. To the writer's knowledge, no entirely satisfactory interpretation has been given to date to account for such findings.

SUMMARY AND CONCLUSIONS

This study was designed to determine (*a*) if significant differences could be found between a group of so-called normal subjects ($N = 29$) and a group of schizophrenic patients ($N = 29$) on five selected level of aspiration variables as measured by the Rotter Level of Aspiration Board and (*b*) if the findings earlier reported by Hausmann regarding the level of aspiration behavior of schizophrenic patients could be reproduced.

The following significant differences served to differentiate between the two groups of subjects: (*a*) the reaction time mean was higher and the variance greater for the schizophrenic group; (*b*) the frequency of high (negative and positive), individual, goal discrepancy score means was greater among the schizophrenic subjects, 34 per cent of the scores obtained by this group falling outside the limits of the normal group; (*c*) the typical response score mean was lower for the schizophrenic patients; (*d*) forgetting to make verbal statements of goals before beginning new trials occurred more frequently among the schizophrenics than among the normal subjects. With respect to the latter finding, there was a greater tendency to forget after unsuccessful than after successful trials among the schizophrenic subjects.

Only one variable, the success score mean, failed to differentiate between the two groups of subjects. However, the interpersonal variability of the schizophrenic group was significantly greater than that of the normal group on this measure.

The differences enumerated above, excluding the differences between variances, were hypothetically accounted for by relating them to the stronger need on the part of the schizophrenic patients to maintain and/or bolster self-esteem. Accordingly, the task in this investigation could be described, interpretively, as one quantifying facets of ego-defense dynamisms.

The assertion made by Hausmann in 1933 that schizophrenic patients, other than paranoid subtype, indiscriminately raise their bids after failure and lower them after success is not supported by the quantitative findings reported in this study. Nor is his second assertion that paranoid schizophrenics cling to high estimates regardless of their achievement scores and in spite of the fact that they were doubly penalized for falling below their estimated scores supported by any findings in this investigation.

REFERENCES

1. BOISEN, A. T. *The exploration of the inner world.* Chicago: Willett, Clark, 1936.
2. DEMBO, TAMARA T. Der Arger als dynamisches Problem. (Untersuchungen zur Handlungs- und Affektpsychologie. X. (Ed. by Kurt Lewin.) *Psychol. Forsch.*, 1931, **15**, 1–44.
3. ESCALONA, SIBYLLE. The effect of success and failure upon the level of aspiration and behavior

in manic-depressive psychoses. *Univer. Ia. Stud. Child. Welf.*, 1940, **16**, 197–302.
4. EYSENCK, H. J. *Dimensions of personality*. London: Kegan, Paul Trench, Trubner, 1947.
5. FESTINGER, L., & KATZ, D. *Research methods in the behavioral sciences*. New York: Dryden Press, 1953.
6. GOLDSTEIN, K. *The organism*. New York: American Book Co., 1939.
7. HAUSMANN, M. F. A test to evaluate some personality traits. *J. gen. Psychol.*, 1933, **9**, 179–189.
8. KLUGMAN, S. F. Emotional stability and level of aspiration. *J. gen. Psychol.*, 1947, **36**, 101–118.
9. MILLER, D. R. Responses of psychiatric patients to threat of failure. *J. abnorm. soc. Psychol.*, 1951, **46**, 378–385.
10. ROTTER, J. B. Level of aspiration as a method of studying personality. II. Development and evaluation of a controlled method. *J. exp. Psychol.*, 1942, **31**, 410–422.
11. ROTTER, J. B. Level of aspiration as a method of studying personality. IV. The analysis of patterns of response. *J. soc. Psychol.*, 1945, **21**, 159–177.
12. WINDER, C. L. On the personality structures of schizophrenics. *J. abnorm. soc. Psychol.*, 1952, **47**, 86–100.

Received Oct. 31, 1953.

ATTITUDE FORMATION: THE DEVELOPMENT OF A COLOR-PREFERENCE RESPONSE THROUGH MEDIATED GENERALIZATION[1]

BERNICE S. EISMAN

Riverside City Schools, Riverside, California

THE present investigation of attitude development has had as its major stimulus Doob's 1947 discussion (7) and reinterpretation of the concept. His placement of attitude within the framework of behavior theory and his emphasis on attitude as a learned response to definable stimuli appears to be a more research oriented and potentially fruitful approach to the problem than has been taken in the past. An equally important source of stimulation for this study has come from the work on mediated and semantic generalization, particularly from the work of Birge (2), Cofer and Foley and associates (3, 4, 5, 8, 9), Keller (11), Razran (13, 14), and Reiss (15, 16).

An attitude is here defined, as suggested by Doob, as a learned implicit anticipatory response having both cue and drive properties. An anticipatory response is one which originally came after a rewarded response but as a result of being most closely associated with the reward moves forward in the behavior sequence. This results in its "anticipatory" occurrence directly to the external stimulus before its original time in the response series. As a response with stimulus properties, an attitude falls into the class of "pure stimulus acts" defined by Hull (10) as responses whose major function is to serve as stimuli for other acts. It follows, therefore, that once such an implicit response-produced stimulus (r-s) has been established, it can be conditioned to evoke overt behavior. In this way an attitude acts as a mediator of responses. It is these responses, mediated by an attitude, which are measured by attitude questionnaires in attitude experiments, etc.

The property of drive value ascribed to an attitude is derived from the fact that an attitudinal response is assumed to produce strong stimulation. Any strong, persistent stimulus is considered by Miller and Dollard (6, p. 30) to constitute a drive. Reduction of the strong stimulus is drive reducing and therefore reinforcing. It is expected, therefore, that any overt response which leads to a significant reduction of the strong stimulus will be learned. Ascribing drive value to attitudes distinguishes them from other implicit, anticipatory response-produced stimuli and is in line with traditional views. Attitudes, wrote Allport (1), "both motivate and guide; they supply both drive and direction." Sherif and Cantril (17, p. 300) offer, as one criterion of an attitude, its affective property. Krech, too, speaks of attitudes as motivational, thus distinguishing them from beliefs (12, p. 152).

STATEMENT OF THE PROBLEM

The purpose of the present study has been to demonstrate experimentally that attitude formation is amenable to laboratory observation and control, and can be adequately conceptualized and explained by reinforcement learning theory. The general hypothesis of this study can be formally stated as follows: A positive preference or attitude toward an object with which an individual has had only neutral experience can be developed through mediated generalization. In more operational terms, it is predicted that if individuals are taught to attach the same label to two stimuli which differ from one another in all other relevant ways, and if these individuals subsequently have rewarding experience with one of these stimuli, then a positive preference for the other stimulus, as well as for stimuli resembling the latter, will be demonstrable. This type of learning experience follows the general paradigm for mediated generalization. In this study the subjects develop a positive response for a color after being subjected to such a series of learning experiences.

To test this general hypothesis four different test situations, in which a color preference could be demonstrated, were employed. These

[1] This paper is based on a portion of a dissertation, prepared under the guidance of Professor F. Fearing, and submitted in partial fulfillment of the requirements for the Ph.D. degree, U.C.L.A., 1953. The author also wishes to thank Professors Gilhousen, Maltzman, Centers, Seagoe, and Turner for their critical suggestions, and Miss C. Seeds, Principal of the University Elementary School, for her cooperation in supplying the subjects for the reported experiment.

situations, differing in context, complexity, and social significance, will be described in detail below.

METHOD

Apparatus and Subjects

A wooden stage, 20.5 in. high, 30 in. wide, and 19 in. deep, with a raised, slanted platform, at an angle of about 35 degrees from the horizontal, was constructed and used for the presentation of various stimulus objects to the subjects. An unbleached muslin curtain could be pulled in front or away from the platform in order to hide or expose objects on it. Three small depressions about 8 in. apart were drilled into the platform to serve as receptacles for marbles which were used as rewards during certain portions of the training.

The following stimulus objects were used: (*a*) three wooden *geometrical blocks*, each painted white and .75 in. thick, triangular, circular, and square, respectively; (*b*) three *thin rectangular blocks*, each 2.5 in. by 6 in. and .75 in. thick, painted green, yellow, and black, respectively; (*c*) three metal *jar tops*, 2.5 in. in diameter, also painted green, yellow, and black, respectively; (*d*) three *glasses*, 3.5 in. high, painted green, yellow, and black, respectively; (*e*) three *thick rectangular blocks*, 1.5 in. by 1.5 in. by 4.5 in., painted green, yellow, and black, respectively; (*f*) three white wooden *"nonsense" blocks*, .75 in. thick, and of three different nonsense shapes.

Eighty-one subjects (*Ss*) were used, 41 boys and 40 girls, ranging in age from 5 years, 10 months to 8 years, 10 months. All *Ss* were attending an elementary school operated by the Department of Education of U.C.L.A. All members of each of the classes used in this study were utilized by the experimenter (*E*), i.e., there was no arbitrary selection of *Ss* from within any class. The *Ss* were told that *E* had a new game she wanted to try out and that she would play with each of the children in a class, one at a time. Owing to the nature of the school, all children in it were accustomed to observation by strangers and many of them were also accustomed to leaving their classes for some time for testing or other "play"-research purposes.

Procedure

The 81 *Ss* were randomly divided into four groups with 9 *Ss* in Group I, 18 *Ss* in Group II, and 27 *Ss* each in Groups III and IV. One third of the *Ss* in each group was randomly selected to be trained positively for each of the three colors, yellow, green, and black. The *E* worked with one *S* at a time. Since the name of each *S* had been assigned to an experimental group prior to the start of the experiment, the order in which *Ss* came to *E* was not fixed, and either the teacher or the last *S* chose the next *S*.

Part 1. Name learning for the geometrical blocks. Each *S* sat facing *E* and the front of the stage. The *Ss* in all four groups were treated alike during this portion of the training. The three geometrical blocks were placed on the stage by *E*, and *S* was told that all three had names which *S* had to discover, the names being "egg" (for the triangle), "car" (for the circle), and "shoe" (for the square). Each *S* was shown one block at a time in a predetermined random order which was the same for all *Ss*. When a block was presented, *S* guessed at its name and *E* told *S* if he was right or wrong. The *S* was given as many trials as was necessary in order for him successfully to identify each of the three blocks five times in succession. Each *S* was told that learning the names of the blocks was necessary in order to proceed to the next part of the game during which he could win some marbles.

Part 2. Reward series. Here, too, *Ss* in all four groups were treated alike. The geometrical blocks were again placed on the platform and *S* was told that *E* would hide a marble under one of them when the curtains were closed and then ask *S* to guess where the marble was. The *S* was instructed to point to the block under which he thought the marble was hidden and to call the block by its *name*, the name learned in Part 1. When placing a marble under the proper block, *E* was careful to make extraneous noise so that sound would not serve as a reliable cue with respect to the hiding place of the marble.

A marble was always hidden under the same block for any given *S*. For those who were to be trained to prefer green, the marble was always placed under the triangle ("egg"); for those to be trained for yellow, the marble was placed under the circle ("car"); and for those to be trained for black, the marble was always placed under the square ("shoe"). The relative positions of the blocks were shifted per trial according to a random, predetermined order which was constant for all *Ss*.

Each time a marble was hidden *S* was permitted only one guess as to its location. If *S* guessed correctly, he kept the marble; if he guessed incorrectly, *E* closed the curtains and hid the marble again, under the same block but in a new position with respect to the other two. The reward series ended when *S* had made six consecutive correct guesses. If the criterion was not reached by the fourteenth trial, a correction procedure was instituted, i.e., when *S* guessed incorrectly, *E* lifted the correct block and showed *S* where the marble was.

The purpose of this part of the training was to provide differential reinforcement for one of the three names ("egg," "car," or "shoe") which *S* had learned in Part 1. This is the only portion of the training in which such differential reinforcement could be obtained.

Part 3. Name learning for the thin rectangular blocks. Immediately after the reward series, *E* presented the three colored rectangular blocks, instructing *S* that one was called "egg" (the green block), one was called "car" (the yellow block), and one was called "shoe" (the black block). The *S* was again asked to discover which block had what name. The procedure followed here was identical to that followed in Part 1 and again *Ss* in all four groups were treated alike.

Part 4. Review series. Groups I and II. Immediately following Part 3, *Ss* in both of these groups were tested for recall of the names of the geometrical blocks and of the colored rectangles and for recall of the marble game (reward series).

Groups III and IV. For *Ss* in these groups a longer and more elaborate review was given in order to introduce more reward and color-naming experience. This was done because the results of a preliminary investigation had led to the conclusion that owing to the

nature of the test situations (to be described below) given *Ss* in Groups III and IV, additional reward and naming experience would be necessary for mediated generalization to be demonstrable.

The *E* presented three new stimuli, the colored jar tops, and asked *S* to guess what their names were. The *S* received a marble each time he correctly named all three stimuli. When *S* reached the criterion of three correct identifications, the colored glasses were presented and the same procedure was followed. The *E* next presented the three thick rectangles and asked *S* to guess their names. Correct identifications were calling the green objects "egg," the yellow ones "car," and the black ones "shoe."

The *Ss* were then reviewed on the marble game (reward series). Then one jar top, one glass, and one rectangle, each of a different color, were presented. All three of the possible combinations were presented successively and *S* was asked to name the objects. No marble rewards were given.

Part 5. Test for mediated generalization and color preference. Here the major differential treatment for the four experimental groups was introduced. The tests were given immediately after the review.

Group I. The *Ss* in this group were presented with the three thin rectangular blocks used in Part 3 and told, "Now we'll play the marble game with these. You tell me where you think the marble is. Don't forget to tell me the name of the block you think the marble is under." With the curtains hiding the platform, *E* placed a marble under each block. Only one trial was given, and after *S* made his choice he was told that there would be no more games. A right choice was picking that colored rectangle which had the same name as the geometrical block under which *S* had consistently found marbles in Part 2.

Group II. The *Ss* in this group were also told that they were going to play the marble game again. The curtains were drawn, but instead of placing the three thin rectangles on the platform, as was done with Group I, *E* substituted three new stimuli, the colored jar tops, with which *Ss* in Group II had had no prior experimental experience. A marble was placed under each jcr top. When the curtains were opened, *S* was told to choose that jar top under which he thought a marble was hidden and to give its name ("egg," "car," "shoe"). The *E* did not tell *Ss* that these were the names of the new objects, and *Ss* had had no experience in so naming them. Only one choice trial was given. A correct choice was picking that jar top which was the same color as the rectangle which *S*, in Part 3, had learned had the same name as the geometrical block under which marbles had consistently been found in Part 2.

Group III. The *Ss* in this group were presented with three new stimuli, the white "nonsense" blocks, and instructed that one was called "green," one "black," and one "yellow." The *S* was told that he had to discover which block had what name. The same name-learning procedure as was used in Parts 1 and 3 was utilized here, and again a criterion of five successive correct identifications of each block was employed. After *S* reached the criterion, he was asked, "If you could take one of these blocks home, which would you pick?"

The color name learned for each of the three "nonsense" blocks was varied from *S* to *S* in order to correct for possible preference for one of the blocks because of its shape. Thus, for example, of three subjects, all trained to prefer black ("shoe") by having been given differential reinforcement for the square block ("shoe") in Part 2, each learned to call a different "nonsense" block "black" in this test situation. Each of the three *Ss* was expected to choose that block which he had learned to call "black."

Group IV. The *E* told *Ss* in this group a story about a pair of twins who had just moved into a new neighborhood. During the course of the story the twins meet three groups of children, a green, a yellow, and a black group. The play activities described for the three groups are almost identical. The climax of the story is reached when the twins decide to have a birthday party and their parents insist that they can invite only *one* of the groups of their new friends. The twins must choose only one of the groups for their party. Each *S* was asked by *E* to choose the group he would invite. A correct choice was picking that group having the same color as the rectangular block which *S* had learned, in Part 3, had the same name as the geometrical block under which he had consistently found marbles in Part 2.

RESULTS

Table 1 summarizes the generalization results obtained from *Ss*' responses in their respective test situations. The numbers in parentheses represent expected frequencies. It was expected that since each *S* had to choose among three colors or three colored objects in the test situation, one third of the group could choose the correct (i.e., trained for) color and two thirds of the group could choose the incorrect color by chance alone. A significant deviation from these proportions in the direction of more than one third *Yeses* and less than two thirds *Noes* refutes the null hypothesis and lends support to the experimental hypothesis.

TABLE 1

GENERALIZATION RESULTS: RESPONSES OF SUBJECTS IN ALL GROUPS IN THEIR RESPECTIVE TEST SITUATIONS

Group	*N*	No. of *Ss* Making Correct Choice	No. of *Ss* Making Incorrect Choice	χ^2	p*
I	9	8 (3)	1 (6)	—	<.002**
II	18	14 (6)	4 (12)	16.00	<.001
III	27	16 (9)	11 (18)	8.17	<.01
IV	27	17 (9)	10 (18)	10.67	<.01

* Probability that the difference between obtained and expected frequencies could have occurred by chance alone.

** Since the theoretical frequencies and total *N* of Group I are too low for the reliable use of χ^2, the probability figure shown was obtained by evaluating the data by means of the binomial expansion (both tails of the distribution considered).

It can be seen from Table 1 that the performance of all groups verified the experimental hypothesis. The null hypothesis of no significant difference between obtained and theoretical frequencies can be safely rejected, and the training given each group is thus shown to be adequate for the generalization of a positive choice response from one stimulus to a previously neutral one.

A chi-square test of independence was also run on the data shown in Table 1 to see if there were any significant differences among the four groups in the success-failure ratio. No such differences were found ($\chi^2 = 3.0$; $p > .30$). A test between Groups I and II combined and Groups III and IV combined also yielded insignificant results, although there is an indication that the more difficult task required of *Ss* in these latter two groups served to lower the number of correct responses ($\chi^2 = 3.37$; $p > .05$).[3]

General Discussion and Conclusions

The results of this experiment have shown that conditions which allow for mediated generalization, and which, in certain cases, offer *Ss* additional opportunities for reward and "labeling" experience, are sufficient to produce a positive response toward a previously neutral color with which *Ss* had had no differential reward experience. From this positive response, a favorable "attitude" towards this color is inferred. In addition, this favorable attitude has been demonstrated in more than one type of situation.

Space limitations do not permit a full presentation of the analysis of the results of this study in terms of learning theory concepts and principles. One illustration of the kind of analysis performed will have to suffice.

Figure 1 represents the conditions for and results of training for *Ss* in Group IV. Stage A represents Parts 1 and 3 of the experimental training, i.e., the name-learning sessions; Stage B represents Part 2, the reward session; and Stage C represents Part 5, the test situation. Stage A might have been better represented by three diagrams—one for each geometrical block and colored rectangle pair to which *S* learned to give the same name. For convenience, only one diagram, however, is used to represent all three. In stage B only one S_g is referred to, depending upon which geometrical block a particular *S* was consistently rewarded for choosing.

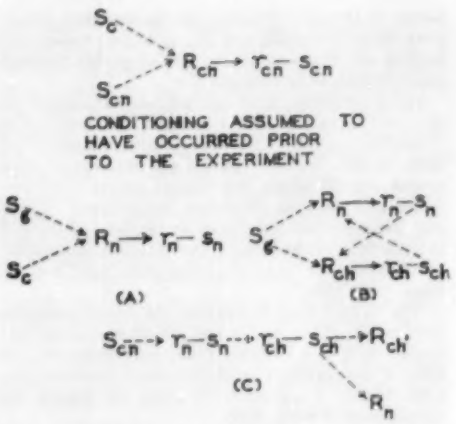

FIG. 1. CONDITIONS FOR AND RESULTS OF TRAINING FOR *Ss* IN GROUP IV

S stands for stimulus of either geometric blocks (g) or colors (c); *R* stands for overt responses of either naming (n) or choosing (ch); r stands for implicit responces and s stands for stimuli produced by these implicit responses.

In stage C, the test situation, color names of three groups of children in a story were presented as oral stimuli to the *Ss* by *E*. S_{cn} represents these color-name stimuli. Each of the color-name stimuli can evoke a specific r_n-s_n because of the relationship shown in Fig. 1 under the heading "conditioning assumed to have occurred prior to the experiment." It is assumed that the *Ss*, outside the experimental situation, have learned to make this same response to both S_c (a visual color stimulus) and S_{cn} (an auditory color stimulus, i.e., a color name). This same response is R_{cn} (naming the color, for example, "yellow"). If we now invoke the principle of mediated generalization we can say that any response subsequently conditioned to S_c will also be evocable by S_{cn}. In stage A such a new response was conditioned to S_g, namely the "egg," "car," or "shoe" response, R_n. As a result, when S_{cn} is presented in stage C, it can evoke the implicit and anticipatory component of R_n, namely r_n-s_n.

[3] The influence of age and sex on generalization in this experimental situation was tested by means of chi square. No significant differences were found to exist between *Ss* above and below the median age of their respective groups with respect to success and failure on the generalization tests. Similarly, no significant difference in the success-failure ratio was found to hold between the sexes.

Once we have thus accounted for the evocation of r_n-s_n, we can see how this can evoke the anticipatory component of R_{ch}, namely r_{ch}-s_{ch}, the "attitude." This stems from the conditioning which took place in stage B and from the anticipatory nature of r_{ch}-s_{ch}. In this situation we see that r_{ch}-s_{ch} evokes, not the old R_{ch} of stage B, but a slightly different choice response ($R_{ch'}$). Both choice responses can be understood as being part of a habit family associated with the same stimulus, r_{ch}-s_{ch}, the "attitude." The evocation of the correct choice response, $R_{ch'}$, cannot be fully explained without assuming, as we have done, its evocation by its implicit component, the attitudinal response-produced stimulus r_{ch}-s_{ch}.

There is one part of our diagram which has thus far been ignored. We have said that in stage *B* r_n-s_n becomes conditioned to R_{ch} and that similarly r_{ch}-s_{ch} becomes conditioned to R_n. That this is indeed the case was demonstrated by the fact that many *S*s, when choosing the colored group of their choice also used an "egg," "car," or "shoe" name to refer to it. Stage C, therefore, has been diagrammed to include the evocation of R_n by r_{ch}-s_{ch}.

Analyses of the conditions and results of training for *S*s in Groups I, II, and III involve the application of similar principles.

It is hoped that the present study represents a step in the direction of experimentally validating certain principles of learning theory on a human level in more or less uniquely human situations. The author has shown that at least one of these principles (mediated generalization) can be used to predict a certain type of human behavior, preference for a class of stimuli with which individuals have had only neutral experience. This type of behavior is common and shows itself most dramatically in the exhibition of attitudes toward social or ethnic groups with which individuals have had little or no direct experience. That such attitudes may be directly taught by parents or other "teachers" and maintained by group reinforcement is one possibility. That they may also be formed in the indirect manner experimentally demonstrated by this study is another. It is not too big a step to assume that if children can learn to prefer the color yellow, for example, over the colors green and black as a result of the type of training given them in this investigation, they can also, as a result of similar training, learn to prefer people called Englishmen over those called Poles and Italians, even though they have had no differential reward or punishment experience with any of those three national groups. The essential requirement for this learning is that individuals should learn to attach the same labels to the word "Englishman" as to stimuli with which they have rewarding experience, and to attach the same labels to the words "Pole" or "Italian" as to stimuli with which they have had neutral or punishing experience.

SUMMARY

Eighty-one elementary school children of grades 1, 2, and 3 were trained to label a white triangular block and a green rectangular block "egg," to label a white square block and a black rectangular block "shoe," and to label a white circular block and a yellow rectangular block "car." In a subsequent situation all *S*s were rewarded for choosing one particular geometrical white block. The *S*s in Group I were then tested for mediated generalization by being asked to choose which of the colored rectangles they thought would bring a reward. The *S*s in Group II were presented with three new, i.e., previously unexperienced, color stimuli and asked to guess which one would bring a reward. The *S*s in Group III learned to label three different white "nonsense" blocks "yellow," "black," and "green," respectively, and then were asked to choose the block they would like to take home. The *S*s in Group IV were told a story about three groups of children, a yellow, a black, and a green group, and then asked which group they would invite to a hypothetical birthday party.

For each group it was predicted that *S*s would choose that stimulus in the test having the same actual color or color name as the colored rectangle which they had learned had the same "egg," "car," or "shoe" name as the geometrical block for the choice of which they had previously been rewarded. These predictions were verified. Significantly more *S*s in each group made the correct choices in the test than would have been expected by chance.

The investigation was designed to test the general hypothesis that a color-preference response (positive attitude) could be developed through mediated generalization, and that this preference could be demonstrated in four situations differing in context, complexity, and

social significance. An analysis, in terms of reinforcement learning theory principles and concepts, of the experimental conditions and results for Group IV was presented as illustrative of the type of analysis necessary for a full explanation of the behavior of *Ss* of all four groups in their respective test situations.

REFERENCES

1. ALLPORT, G. W. Attitudes. In C. Murchison (Ed.), *A handbook of social psychology*. Worcester, Mass.: Clark Univer. Press, 1935. Pp. 798–844.
2. BIRGE, JANE S. The role of verbal responses in transfer. Unpublished doctor's dissertation, Yale Univer., 1941.
3. COFER, C. N., & FOLEY, J. P., JR. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
4. COFER, C. N., & FOLEY, J. P., JR. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.
5. COFER, C. N., JANIS, MARJORIE G., & ROWELL, MARY M. Mediated generalization and the interpretation of verbal behavior: III. Experimental study of antonym gradients. *J. exp. Psychol.*, 1943, **32**, 266–269.
6. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
7. DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.
8. FOLEY, J. P., JR., & MATHEWS, M. A. Mediated generalization and the interpretation of verbal behavior: IV. Experimental study of the development of interlinguistic synonym gradients. *J. exp. Psychol.*, 1943, **33**, 188–200.
9. FOLEY, J. P., JR., & MACMILLAN, Z. L. Mediated generalization and the interpretation of verbal behavior: V. Free association as related to differences in professional training. *J. exp. Psychol.*, 1943, **33**, 299–310.
10. HULL, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **6**, 511–525.
11. KELLER, MARGARET. Mediated generalization: the generalization of a conditioned GSR established to a pictured object. *Amer. J. Psychol.*, 1943, **56**, 438–448.
12. KRECH, D., & Crutchfield, R. S. *Theories and problems of social psychology*. New York: McGraw-Hill, 1948.
13. RAZRAN, G. H. S. Salivating and thinking in different languages. *J. Psychol.*, 1935–36, **1**, 145–151.
14. RAZRAN, G. H. S. A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science*, 1939, **90**, 89–90.
15. REISS, B. F. Semantic conditioning involving the GSR. *J. exp. Psychol.*, 1940, **26**, 238–240.
16. REISS, B. F. Genetic changes in semantic conditioning. *J. exp. Psychol.*, 1946, **36**, 143–152.
17. SHERIF, M., & CANTRIL, H. The psychology of attitudes. *Psychol. Rev.*, 1945, **52**, 293–319.

Received December 4, 1953.

THE RELATIONSHIP BETWEEN SYMBOLIC AND MANIFEST PROJECTIONS OF SEXUALITY WITH SOME INCIDENTAL CORRELATES

RUSSELL A. CLARK AND MINDA RAE SENSIBAR

University of Michigan *Harvard University*

THIS paper will be concerned with a further analysis of the data collected from two of the experiments (experiments A and C) described in greater detail in an earlier article by Clark (2). In brief, experiment A consisted of collecting TAT stories from one group of subjects (*S*s) after they had been exposed to photographic slides of attractive nude females. The control group for this experiment consisted of comparable *S*s who wrote stories after having been exposed to slides of a nonsexual nature. Experiment C was essentially a repetition of experiment A with the addition that the TAT stories were written under the influence of alcohol in a beer-party setting. That is, the experimental group took the TAT during a fraternity beer party after having been exposed to slides of attractive nude females. The control group in this experiment took the TAT during a beer party without prior exposure to such slides. Hereafter experiment A will be referred to as the nonalcoholic condition and experiment C as the alcoholic condition.

The TAT stories were scored for the presence of manifest sexual content. This sexual content was partitioned into three components: (*a*) primary—explicit or implicit evidence for sexual intercourse; (*b*) secondary—evidence for the occurrence of such secondary sex activity as kissing, dancing, fondling, etc.; (*c*) tertiary—characters in the stories perceived as sweethearts, on a date courting, in love, etc., but not engaged in either primary or secondary sexual activity. Sexual activity was scored only once for each story, and that which was biologically most sexual received priority.

The analysis of the data from the nonalcoholic condition revealed that the sexually aroused group expressed significantly *less* manifest sex content in their stories than did the unaroused control group. The analysis of the data from the alcoholic condition showed that the aroused group expressed significantly *more* manifest sex than did their unaroused controls. These results along with others described in the previous article led to the conclusion that under normal conditions increased anxiety or guilt accompanying the increased sexual arousal accounts for the inhibition of the manifest sex content. However, under conditions involving alcohol and the permissive beer-party setting the guilt or anxiety over expressing manifest sex is so reduced that the aroused group expresses significantly more sex than does the control.

PROBLEM AND PROCEDURE

With these foregoing results in hand a Freudian hypothesis immediately came to mind. When the manifest expression of sexuality is being inhibited, as seems to be the case for the aroused group under nonalcoholic conditions, perhaps a disguised form of expression such as symbolism will be exhibited. Freud's discussion of this situation with respect to dreams is too well known to require any elaboration here. Therefore, it was decided to score for the presence of sexual symbolism. This presented certain difficulties. The major difficulty was encountered when we tried to score merely for the presence of all separate objects that are considered to be potential symbols. The reason for our difficulty was that the TAT pictures which were employed clearly depicted objects that are considered classical sexual symbols. For example, picture 12BG clearly shows a boat resting in a pond under a tree in a general wood setting. The boat, tree, and rolling landscape are sexual symbols mentioned by Freud and others. Picture 14 shows the silhouette of a boy in a window. A window or portal is, of course, another sexual symbol. Other such examples could be cited. Thus, we were getting large numbers of symbols that were merely mentions of objects that were clearly depicted in the pictures. This did not seem to be a sensible procedure. Our final decision was not to score for the presence of symbolism if there was only the mere mention of a given classical symbol that was clearly depicted in the TAT picture. In fact, we did *not* score the mere mentioning of a classical symbol even if it was not clearly depicted in the picture. In order to be scored for symbolism a story had to meet the following criterion: The classical symbol or symbols mentioned had to be utilized or involved in some action which in and of itself could be interpreted as being symbolic. That is, isolated objects which fulfilled the requirements for sexual symbols were scored only if they appeared in a symbolic context that could be interpreted as signifying intercourse, masturbation, tumescence-detumescence cycles, etc. For example, if in response to picture 14 a person wrote .hat this was a boy looking out of his dormitory window, this was *not* scored as symbolic, but if it was a thief *climbing up* to go through the *window* in order to steal the *jewels* in the *house*, this *was* scored as symbolic. In this latter statement the window, which is a classical symbol and clearly represented in the picture, is also a part of a

classically symbolic action sequence, i.e., climbing, breaking into house, stealing jewels. In essence what we scored for was not the mere presence of isolated objects that could or could not be symbols, but for themata that could be interpreted in a symbolic fashion. These themata, of course, involved objects which are considered to be the symbols for the male or female genitals, breasts, etc.

In scoring, the stories were given weights of two, one, or zero. A weight of two indicated that symbolism was strongly present in the story, a weight of one indicated that some symbolism was present, and zero indicated that no symbolism could be detected. An individual's total score was the sum of the weights given his five separate stories. The scoring was done independently by two scorers after they had familiarized themselves with the literature on sexual symbolism. The product-moment correlation between the initial two sets of total scores for both the alcoholic and nonalcoholic conditions was +.82. The scoring used for the present analysis represents a joint scoring in which differences in the two sets of scores were reconciled.

RESULTS AND DISCUSSION

A t test of the difference between the two groups of the nonalcoholic condition (cf. also Table 1) demonstrated that the aroused group exhibited a significantly higher symbolic sex score than did the control group ($t = 2.19$, $p < .05$). It will be recalled, of course, that this aroused group showed significantly less manifest sexual content than its control group. This finding, then, is in line with Freud's general hypothesis that anxiety can cause manifest sex content to be inhibited, but that sexual motivation can find expression in symbolic form. Also if Freud's hypothesis is correct, one would expect the groups in the alcoholic condition to express little symbolism because of the fact that once anxiety was reduced by the consumption of alcohol and the permissive party atmosphere most of the sexual expression would be manifest in nature. This, in fact, turned out to be the case. In this condition both the control and the aroused group expressed very little symbolism and in almost identical amounts. The average symbolism score for the combined groups in this condition was .45. This value is lower than that obtained for either the control group ($t = 1.79$, $p < .08$) or the aroused group ($t = 3.87$, $p < .01$) of the nonalcoholic condition. These findings, of course, are in complete accord with the classical Freudian contention. That is, if anxiety is present, the libido finds expression in a disguised or symbolic fashion. When anxiety is absent, sexual expression is channeled directly in manifest form.

A recent paper by Hall (6), however, takes exception to this Freudian contention. In general Hall maintains that sexual symbolism in dreams is not a disguised expression of sex, the purpose of the disguise being to smuggle the content past the censor, but rather a means of representing as clearly as possible the particular conception of sexuality that the dreamer has in mind. One of several cogent reasons that Hall has for offering his alternative theory is that in his collection of dreams he often found both manifest and symbolic expression of sexual activity in the same dream or same sequence of dreams. The question that Hall raises concerns the efficacy of disguising the sexual content if at the next moment it is revealed in manifest form.

Hall's theory was of particular interest because the same phenomenon was noted in this study in the TAT stories. That is, an individual would give both strong manifest and symbolic sex in the same story or series of stories. To throw additional light on this somewhat paradoxical question we decided to examine the intragroup relationship between symbolic and manifest expression of sex. If Freud's contention is correct, you would again expect a negative relationship between the two as was the case with the intergroup comparisons. If Hall's formulation is correct, one might possibly expect even a positive relationship. It was feasible to do this only for the two groups of the nonalcoholic condition. The frequency of appearance of symbolism was so very low under conditions of alcohol that any kind of intragroup trend analysis was meaningless. The obtained relationship for the nonalcoholic condition can be seen in Table 1. Table 1 presents the results of an analysis of variance in which both the aroused and control groups were divided into thirds as equally as possible based on the *manifest* sex score. This manifest sex score was derived by giving the weight of two to a primary sex story, a weight of one to a secondary or tertiary sex story, and a weight of zero to a story containing no manifest sex. The breakdown into thirds of the distribution was made separately for the control and aroused groups. That is, the "highs" in the aroused group represent those with the highest manifest sex content

TABLE 1

MEAN SCORE ON SEXUAL SYMBOLISM AS A FUNCTION OF EXPERIMENTAL CONDITION AND MANIFEST SEX SCORE (NONALCOHOLIC CONDITION)*

Experimental Condition	High Manifest Sex Group		Moderate Manifest Sex Group		Low Manifest Sex Group	
	N	Mean	*N*	Mean	*N*	Mean
Aroused	15	1.87	14	.64	10	1.70
Control	9	.89	16	.56	13	1.00

Analysis of Variance†

Source	Sums of Squares	*df*	Variance Estimate
Total	38.20	76	
Experimental treatment	2.15	1	2.15
Manifest sex score	3.18	2	1.59
Error	32.87	73	.45
Experimental treatment × Manifest sex score	0.51	2	.26
Within cells	32.36	71	.46

* In the statistical treatment of the data in this table the means are based on the raw data. However, the actual *t* tests or *F* tests were based on transformed data using the square-root technique. This was done because the variance estimates were not homogeneous and seemed to be proportional to the means. Also, in the analysis of variance adjustment was made for the disproportionate subclass frequencies. This was done according to the method offered by Johnson (7).

† $F = \frac{\text{Experimental treatment}}{\text{Error}} = 4.78 \quad p < .05;$

$F = \frac{\text{Manifest sex score}}{\text{Error}} = 3.53 \quad p < .05.$

in their own group, but they have significantly lower manifest sex scores than the highs in the control group. Lastly, the entries in the cells represent averages based on the total symbolic sex score previously described.

It can be seen from Table 1 that the relationship between manifest and symbolic sex is neither positively or negatively linear but rather a curvilinear one with those individuals who write either strong manifest sex stories or very weak manifest sex stories exhibiting the greatest amount of symbolism. That this trend is probably not a random one is evidenced by the *F* test involving the variation due to the different levels of manifest sex score. Also the fact that the interaction between the experimental treatment and manifest sex score is insignificant indicates that there is no evidence that the trends in the two groups are dissimilar.

This curvilinear relationship is rather a curious finding. The large amount of symbolism in the high manfest group is, of course, the major deviant finding and in need of explanation. This follows because according to a Freudian formulation the writing of strong sex stories would seem to preclude the possibility of the presence of much anxiety. There are a number of ways in which to account for our finding. One way is to assume that if sexual motivation becomes sufficiently high, there is the tendency for symbolism to be expressed without regard to the presence of anxiety. With this added assumption an interpretation which is in line with Hall's general formulation could also account for the increased amount of symbolism in the aroused group.

One other possibility is, of course, that both Hall and Freud are correct. It may well be that symbolism is used as an alternative means of expressing sex even in the absence of anxiety but becomes more preferred to the extent that anxiety is involved. However, an examination of the data from the alcoholic condition should help clarify this issue. The aroused group from this experiment represents a sexually aroused group presumably with much of the anxiety reduced by the alcohol and permissive party atmosphere. This group gave a large number of manifest sex stories. If Freud's formulation is correct, you would *not* expect much symbolism in this group because of the presumably low anxiety level. If the formulation which is more in line with Hall's theorizing is correct, you would expect a large amount of symbolism. As indicated previously from the comparisons involving the combined groups from the alcoholic condition this is not the case. The average symbolism score for the aroused group alone was .41, which is lower than the aroused group ($t = 3.88$, $p < .01$) and control group ($t = 1.83$, $p = .07$) of the nonalcoholic condition.

It would thus seem that to the extent that anxiety is lacking, symbolism is lacking, and that Freud's formulation is supported. However, there still remains the apparently paradoxical fact that in the intragroup breakdown for the nonalcoholic condition those with high manifest scores also have high symbolism scores. This seems paradoxical because, as mentioned before, the writing of sex stories, especially of a primary nature, would seem to preclude the possibility of the presence of

anxiety. However, there is one possibility that should not be ignored, and this involves the *response-produced* guilt that stems from writing stories of a sexual nature. The individuals who write strong sex stories must normally have a fairly low anxiety level which permits them to do this, but writing sexual stories, especially if primary in nature, may cue off enough guilt to channel the expression into symbolic form. This certainly did seem to be what was happening in some cases in which a primary manifest sex story was followed immediately by one containing symbolism.

The solution to this very paradox has been handled by Miller (10) in slightly different terms in his treatment of approach-avoidance conflicts. Miller points out, "Similar paradoxical effects may be deduced when the attractiveness of the goal is held constant and its repulsiveness varied within the limits allowing the two gradients to cross. As the strength of avoidance at the goal is weakened, the subject will be expected to move forward. But as he moves forward, the strength of approach increases so that stronger avoidance must be aroused before his advance is stopped. The fact that the subject moves nearer to the dangerous goal more than compensates for its reduction in unattractiveness" (p. 440).

In this instance going closer to the goal would involve writing stories of a more primary nature. Therefore, it would be individuals with the initial low gradients of avoidance who would write the most sexual stories and thus be apt to experience the most *response-produced* guilt.

Therefore, what may possibly be happening is something like this. The individuals in the *low* manifest group (both aroused and controls) through past training are highly anxious about expressing sex. For the aroused *S*s this anxiety is, of course, reinforced by the stimulation of the nude slides to which they were exposed. These individuals, therefore, would tend to express most of their sexuality in symbolic terms. The individuals in the *high* manifest group through past training have acquired less anxiety over sexuality and therefore approach close enough to the goal to write stories of a primary nature which in turn cues off quite a bit of guilt resulting in symbolic expression of sexuality. Now what about the individuals in the moderate manifest category? They could be individuals with a moderate amount of anxiety who write stories with mild sexuality in them, but this is not enough to cue off much guilt, so they have only a moderate amount of symbolism in their stories.[1]

There still remains at least one apparent inconsistency: Why is it that in the intragroup comparisons the individuals with the highest manifest sex scores give high symbolism, whereas in the intergroup comparisons just the reverse holds true? Why should not the control group of the nonalcoholic condition give as much or more symbolism than the aroused group in view of the fact that they give significantly more manifest sex? That is to say, is it not possible that the increased response-produced guilt of the control group is enough to offset the stimulus-produced (nude slides) anxiety of the aroused group? One possible answer to this is that in the aroused group nearly *everyone* should have experienced a certain high level of anxiety as a function of examining the slides of attractive nude females, whereas, in the control group perhaps an equivalent level of guilt was reached by only those individuals who wrote stories of a primary manifest nature.

Perhaps an analogy involving the rat in a maze would simplify the entire picture. Assume that hungry rats have in the past been both shocked and fed in the goal box of a straight maze. If one group of rats is put in a maze near the goal box where the fear is great, they will start making avoidance responses. Also assume that in the past the rats have been fed in side alleys, that they have not received very much food there but also little

[1] Although the foregoing analysis makes pretty good sense, it should be obvious that this is an oversimplification. A tacit assumption has been that although individuals varied in the amount of anxiety over sex (avoidance motive), their sex drives (approach motives) were relatively the same. This is probably not a warranted assumption even though it is likely that the variability in anxiety is far greater than the variability in the sex drive. If one allows for variability in the approach motive, it might be that the individuals who wrote primary stories, for example, did so *not* because of an initially lower avoidance motive, but because of a high approach motive. This complicates even further an already highly complicated picture. We regard the *ad hoc* interpretations presented above as mainly serving the function of highlighting the necessity of considering the various pertinent factors. To do definitive work, ideally one would have to have independent measures of the approach tendency, both the stimulus-produced and response-produced avoidance tendencies, plus a knowledge of how these three interacted to produce both manifest and symbolic expressions of sexuality.

or no shock. So in addition to retreating down the alley they may also duck into a side alley and get some partial gratification without anxiety. The other group of rats is placed in the far end of the maze where fear is low. These rats will start to make approach responses. Some of these rats, as Miller (10) points out, will "overshoot" into the territory where the anxiety is high enough to cause them to beat a retreat or duck into the side alleys. However, on the average one would expect both more avoidance behavior and substitute gratification on the part of the first group than the second.

This is analogous to what might be taking place in the sex experiment with sex and anxiety replacing hunger and fear. The side alleys would correspond, of course, to the symbolic expression of sexuality. Avoidance behavior would be comparable to expressing less anxiety-laden sexuality or no sexuality.

Although much of the foregoing has been *ad hoc* conjecture, a few tentative conclusions seem justifiable: (*a*) It does seem that classical symbolism serves as a disguised expression of sexuality. (*b*) Symbolism as a mode of expression seems to be preferred only to the extent that anxiety over sex is present. (*c*) The amount of symbolism in a story is not necessarily a direct measure of the "normal" level of anxiety. For example, if an individual's normal level is quite low, he may approach so close to the goal that this engenders a high level of response-produced guilt which may in turn produce symbolic expression of sexuality. (*d*) The above is a possible explanation of the seemingly paradoxical fact that both strong manifest and symbolic expressions of sexuality appear in the same fantasy sequences.

Incidental Correlates of the Inhibited Expression of Sexuality

In view of the evidence of the inhibited sexual expression obtained under nonalcoholic conditions it seemed worth while to examine the data from this condition to see if a similar finding held for the perception of the sex of the characters depicted in the TAT pictures. A second relationship was investigated because of the vast evidence for a linkage between sex and aggression (see next section).

Perception. Three of the five pictures used in the nonalcoholic condition were sufficiently ambiguous to make possible a misinterpretation of the sex of the characters. Picture 14 from Murray's 1943 series shows the silhouette of a figure standing at a window. This figure is usually perceived as a male by male *S*s but it is sometimes perceived as a female. Picture 20 from this same series depicts a shadowy figure standing under a lamp post. This again is more often perceived as a male than as a female. Picture 10 is also somewhat indistinct and is usually described as a man and a woman in an embrace or confronting one another, but sometimes it is perceived as two men or two women. The contribution of picture 10 to the results was negligible because it was misperceived so rarely, but to be consistent it was counted in the analysis if anyone described the two figures as female. Table 2 demonstrates that the aroused group perceived all of these characters as female significantly more often than did the control group.

TABLE 2

PERCENTAGE OF SUBJECTS PERCEIVING AT LEAST ONE OUT OF THREE AMBIGUOUS FIGURES AS FEMALE

Condition	*N*	Percentage Perceiving at Least One Female	Chi Square
Aroused	39	64.1	6.87*
Control	38	34.2	

* Significant at the .01 level.

This actually is not what one might expect. Bruner and Postman (1), for example, discuss two general effects of emotionality, value, etc. on perception. One is a *defense* process in which the individual fails or is slow to perceive an anxiety-laden stimulus. The other is a *sensitization* process in which the stimulus is perceived more readily. In light of the previous findings which show the aroused group being defensive by expressing sex symbolically rather than manifestly, one might expect the defensive process to be dominant and operate such that ambiguous figures would *not* be perceived as female. This was not the case. However, in a further analysis it was found that for the aroused group in only 23.1 per cent of the stories in which misperceptions occurred was the female involved in any kind of manifest sex activity. The same percentage for the control group was 50.0. Thus, it would appear that sensitization is occurring at the perceptual level, but that defensive processes intervene to prevent the involvement of the female figure in any kind of sexual activity.

Aggression. Throughout the psychological literature there appear scattered conjectures and evidence that sexuality and aggression are somehow intimately linked. Freud writes, "That cruelty and the sexual instinct are most intimately connected is beyond doubt taught by the history of civilization, but in the explanation of this connection no one has gone beyond the accentuation of the aggressive factors of the libido" (5, p. 570). Dollard *et al.* (3) report a study by Sollenberger in which there was shown to be a very marked relationship between aggressive behavior and male hormone content of adolescent boys. Ford and Beach (4) cite numerous examples of aggressive behavior occurring during copulation up and down the phylogenetic scale. In a recent study Lindner (8) found that sexual offenders showed more aggression on a projective test than did a control group of nonsexual offenders.

A similar finding exists with the present data. The stories for both groups were scored for the presence of overt aggression. There were five general categories that constituted the scoring system for aggression: (*a*) all occurrences of death except by suicide;[2] (*b*) physical violence—beatings, auto accidents, combat, etc.; (*c*) personal tragedies and misfortunes—homes burning to the ground, loved ones ill, loss of life savings, etc.; (*d*) crimes—theft, breaking and entering, embezzlement, etc.; (*e*) quarrels—disagreements in which anger, hatred, resentment, tension, etc. are involved. Punishment or imprisonment for crimes committed was *not* considered as aggression. The crimes were scored, of course, but not any attendant punishment. Finally, only aggressive acts directed against persons were considered in this present analysis.[3] The scoring was done independently by two separate scorers. A product-moment reliability coefficient was calculated using as the score for each individual the total frequency of appearance of aggressive incidents in his five stories. This correlation was .91.

Table 3 shows that the aroused group expresses significantly less aggression than the control group. It will be recalled that the aroused group under this condition (nonalcoholic) also expressed significantly less manifest sex. This would again seem to show a close linkage between sex and aggression. However, it is not at all clear what the exact nature of this linkage is. If there had been *more* aggression in the aroused group instead of less, this would have been consonant with the frustration-aggression hypothesis expounded by such writers as Dollard *et al.* (3). That is to say, the aroused group can be considered to have been frustrated in that they were aroused but had no opportunity for release. A distinct possibility is that there were greater aggressive tendencies in the aroused group, but these suffered the same fate for the same reason as the sexual tendencies discussed in the previous sections. That is to say, it might easily be that the expression of any frustration-produced aggression would be inhibited by the presence of anxiety over expressing aggression. It might be that some measure of a disguised or substitute form of aggression would show a higher frequency in the aroused group as was the case with expression of sexuality.

TABLE 3

PERCENTAGE OF SUBJECTS EXHIBITING AT LEAST THREE INSTANCES OF OVERT AGGRESSION

Condition	*N*	Percentage Giving at Least Three Aggressive Responses	Chi Square*
Aroused	39	10.2	17.57**
Control	38	57.9	

* Corrected for continuity.
** Significant at the .01 level.

Although it is a little difficult to conceive of something like symbolic aggression, there is a somewhat different approach that is pertinent. McClelland (9) in his discussion of aggression cites evidence that if there is strong anxiety present, aggressive tendencies appearing in fantasy will be projected into impersonal, vague, or unspecified agents. That is, a person is killed not at the hands of another person but by an animal, through sickness, accident, etc.

Following this lead all instances of aggression appearing in the stories were scored as to whether the agent of the aggression was personal in source or whether the agent was

[2] For the purposes of the present analysis and a later analysis, it was considered expedient to omit suicides because their theoretical significance is quite complex. A check was made, however, and it was found that their inclusion would not have altered the nature of the findings.

[3] Those *Ss* who had no aggression in any of their stories could not, of course, be included in this analysis.

impersonal or unspecified. In most cases this was a straightforward judgment made with little hesitancy. However, two general categories were somewhat equivocal. The first involved auto accidents, railroad accidents, etc. It could be argued that because there must be a person behind the wheel this type of aggression should be considered, in the last analysis, to be personal in source. Nevertheless, because in this general situation intent to harm is absent, and because the emphasis is definitely on the vehicle as the instrument of harm, we felt strongly that this situation belonged in the impersonal category. The second situation which was not completely clear involved mere statements that an individual had been killed in the war.[4] These statements are to an extent depersonalized, but because war usually involves interpersonal conflict with intent to harm, these instances of aggression were scored as personal. The scoring was again done independently by two individuals and product-moment reliability coefficients were calculated. An individual's score consisted of the total number of personal or impersonal aggressions. Separate reliabilities were calculated for the two types of aggression. For aggression which was personal in source the correlation was .93; for impersonal aggression the correlation was .88.

The results of this analysis showed that for the control group 30.2 per cent of the aggression was impersonal or unspecified whereas the aroused group exhibited 53.6 per cent as impersonal or unspecified. The analysis in Table 4 demonstrates that this is a reliable difference. Therefore, although the aroused *S*s express significantly less aggression than the control group, an appreciably greater percentage of their aggression is impersonal in source. Thus, a reasonable conclusion would seem to be that the aroused *S*s were not only anxious about expressing manifest sex, but also anxious about expressing manifest aggression; and when they did express manifest aggression, there was a tendency to project it into impersonal or unspecified sources.

The possibility that *displaced* aggression might show the above trend was entertained,

[4] There may be some doubt as to whether "deaths in war" should be considered as personal or impersonal aggression. Analyses omitting these relatively few instances altogether or scoring them as impersonal in source still yield results significant beyond the .05 level of confidence.

TABLE 4

PERCENTAGE OF SUBJECTS GIVING AT LEAST 50 PER CENT OF THEIR AGGRESSIVE RESPONSES AS IMPERSONAL IN SOURCE

Condition	*N*	Percentage Giving at Least 50% Impersonal Aggressive Responses	Chi Square
Aroused	30	66.6	6.70*
Control	35	34.3	

* Significant at the .01 level.

but it is difficult to decide what figure should be the primary target of the aggression from which displacement could be measured. We did score for aggression that was displaced to animals or inanimate objects. There were more of these in the aroused group, but the frequency of occurrence was so low that statistical analysis was precluded.

SUMMARY

This study involved scoring TAT stories for the presence of sexual symbolism. The stories scored came from two different experiments. In one experiment the experimental group was sexually stimulated before taking the TAT, while the control group received no such prior stimulation. The second experiment was conducted in the same fashion except that it was carried out in a beer-party atmosphere under the influence of alcohol. The results showed that under nonalcoholic conditions the aroused group expressed significantly more symbolism than the control group. It had been shown previously that this aroused group inhibited the manifest expression of sex. Under conditions of alcohol both groups gave very little symbolism. It had previously been shown that these two groups expressed large amounts of manifest sexuality. These results are, of course, in line with the Freudian hypothesis that inhibited sexuality finds an outlet symbolically in fantasy. However, a second, intragroup, analysis showing the amount of symbolism, expressed as a function of the amount of manifest sex expressed, revealed the relationship to be curvilinear with both low and high manifest sex corresponding to high symbolism. Various interpretations are discussed for this latter finding. A further incidental finding revealed that the aroused group perceived ambiguous figures in the TAT as female more often than did the control.

However, a smaller percentage of these "misperceived" females became engaged in sexual activity in the aroused group than in the control group. Another finding dealt with the expression of manifest aggression. The aroused group which expressed significantly less manifest sex in their stories also expressed significantly less manifest aggression. It was also found that the aggression which did appear in the stories of the aroused group was more apt to be attributed to impersonal or unspecified sources.

REFERENCES

1. BRUNER, J. S., & POSTMAN, L. Emotional selectivity in perception and reaction. *J. Pers.*, 1947, **16**, 69–77.
2. CLARK, R. A. The projective measurement of experimentally induced levels of sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391–399.
3. DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.
4. FORD, C. S., & BEACH, F. A. *Patterns of sexual behavior*. New York: Harper, 1951.
5. FREUD, S. *The basic writings of Sigmund Freud*. New York: Random House, 1938.
6. HALL, C. S. A cognitive theory of dream symbols. *J. gen. Psychol.*, 1953, **48**, 169–186.
7. JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.
8. LINDNER, H. Sexual responsiveness to perceptual tests in a group of sexual offenders. *J. Pers.*, 1953, **21**, 364–374.
9. MCCLELLAND, D. C. *Personality*. New York: William Sloane, 1951.
10. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 431–465.

Received January 13, 1954.

TEST ANXIETY AND BLAME ASSIGNMENT IN A FAILURE SITUATION

JOHN DORIS AND SEYMOUR SARASON

Yale University

PREVIOUS studies of the test-anxiety groups suggested that the high-anxiety and low-anxiety groups might differ in their preferential use of defense mechanisms (1, 6). Therefore, it was decided to determine whether or not HA and LA groups would differ in their assignment of blame in a failure situation, i.e., upon experiencing failure in a test situation in which opportunity is provided for assigning blame to *self* or *other-than-self* would HA and LA subjects differ in their assignment of blame?

The present study is concerned with two aspects of blame assignment: (*a*) Does the direction of blame assignment in a failure testing situation vary with the level of anxiety of the testee? (*b*) In the case of repeated failure in a testing situation, when each failure offers a fresh opportunity for blame assignment, will the variability of the direction of blame assignment differ with the level of anxiety of the testee, i.e., with repeated failure will the groups differ in the constancy with which they blame the *self* or *other-than-self*?

The prediction was made that the high-anxiety (HA) subjects (*S*s) would blame themselves more in a failure situation than the low-anxiety (LA) *S*s, a conclusion derived from a consideration of the greater propensity for self-relevant responses in those with HA. In previous studies (3, 5), those responses which were self-relevant—in contrast to the task-relevant responses—were considered to include feelings of inadequacy, helplessness, heightened somatic reactions, anticipations of punishment or loss of status and esteem, and implicit attempts at leaving the test situation. Such self-relevant responses are more likely to evoke expressions of self-blame rather than other-than-self blame.

As regards constancy in the direction of blame assignment, a previous study (3) has suggested that fluctuations in the anxiety level of the subjects during the test situation may vary the propensity for self-relevant responses; this in turn would cause variations in self-blame. Presumably, with LA subjects the anxiety level remains uniformly low throughout the test situation. Therefore, with LA the propensity for self-blame should remain relatively constant. In the case of the HA subject, the anxiety level—and its derived propensity for self-relevant responses—might well be subject to change. As the HA subject becomes adjusted to the test situation his initially high anxiety level may drop off, or as he meets repeated failures it may go even higher, or both factors might be operative within the same testing situation with resulting fluctuations in anxiety level. Whichever one of these three conditions obtains in our experiment—a uniform drop, a uniform rise, or a fluctuation in anxiety level—it is obvious that the propensity for self-blame should vary accordingly and hence make for variability in the constancy with which blame is directed toward the self in the HA group.

METHOD

The results presented here are derived from an experiment with 28 HA and 28 LA *S*s selected on the basis of a test-anxiety questionnaire as in previous studies (3, 4, 5). The *S*s were not aware that their selection was based upon their responses to the test-anxiety questionnaire. At the beginning of the testing situation *S* was told that he was going to be given some intelligence tests as part of a study related to the aptitude tests given to every Yale freshman. He was told that his scores in the session would be compared with his aptitude test scores in order to make any necessary adjustments in the latter.

The *S* was then presented with eight performance items taken from various standard intelligence scales:

Vigotsky (*Failed*)
Cube construction #3 from the Grace Arthur (*Passed*)
Picture Completion #1 to #13 from the Wechsler-Bellevue (*Passed*)
Stencil Design #17 from the Grace Arthur (*Failed*)
Object Assembly, Mannikin and Face, from the Wechsler-Bellevue (*Passed*)
Kohs Block Design #16 (*Failed*)
Ferguson Form Board #37006 (*Failed*)
Digit Symbol from the Wechsler-Bellevue (*Passed*)

The situation was so arranged, either by manipulation of the time limits or reference to purported norms for a college population, that *S* would pass four and fail four of the tests. In the above list the failed and passed tests are indicated. The tests were administered by two methods. In Method I they were administered in the order listed. In Method II they were administered in reverse order, beginning with Digit Symbol and

ending with Vigotsky. The *Ss* in HA and LA were randomly assigned to the two methods, so that 14 members of each group were tested by each method. After success or failure on a test *S* was asked to rank in order of their relevance some items that might have contributed to his success or failure as the case might be. He was told that these rankings were necessary in order to facilitate our interpretation of his performance on these tests when we made our adjustment of his scores on the Yale aptitude tests. For failed tests he ranked the following items:

The reasons I failed to complete this task are:

A. The examiner did not make the instructions as clear as they could be and/or his attitude and behavior did not induce me to make my best possible effort. (*Other-than-self blame*)
B. I had some difficulty in understanding the instructions. (*Self-blame*)
C. I did not try as hard as I could because the test is not very interesting or challenging. (*Other-than-self blame*)
D. The task was a very difficult one. (*Neutral*)
E. Nervousness or anxiety interfered with my performance. (*Self-blame*)
F. I don't know why I failed. (*Neutral*)
G. The presence of the examiner (i.e., I work my best when I am not observed). (*Other-than-self blame*)
H. I made several unnecessary blunders which I should not have made. (*Self-blame*)

The *S* was told to assign numbers from one to eight in order of relative importance from most to least. These items were selected by the experimenters with two considerations in mind. (*a*) We wished to have a number of items that would express blame for failure either toward the self or toward the examiner and the testing situation. (*b*) We wished to have items that would cover the more frequent reasons that a testee might advance in explanation of failure on a test. This latter criterion was necessary in order that the subject would not become aware that the examiner was solely interested in finding out where the subject assigned blame for failure. Thus, of the eight items ranked only six were considered to be concerned with assignment of blame (see list). The other two were included because of the frequency with which they are used in explanation of failure. For the tests which were passed there was a parallel set of eight items, three of which assigned credit for success to the self and three which assigned credit to the examiner or test situation. At the end of the session *S* was disabused and told that the tests as administered were not a valid measure of intelligence. Introspective reports were then gathered on *S*'s attitudes during the test situation.

RESULTS

In the analysis of the data the neutral items were discarded and the remaining items reranked in the order determined by *S*'s original ranking. These remaining items consisted of the following groups: self-blame items and other-than-self blame items for the failed tests, and self-credit and other-than-self credit items for the passed tests. The mean ranks assigned to the self-blame, other-than-self blame, self-credit, and other-than-self credit items were then obtained for each subject. These mean ranks were then considered as the *S*'s scores for self-blame, other-than-self blame, self-credit, and other-than-self credit. Since an *S*'s self-blame score automatically determines his other-than-self blame score, our statistical treatment treats only the self-blame score. Similarly, for the success tests only the self-credit scores were considered.

TABLE 1

MEAN SELF-BLAME SCORES FOR GROUPS WITH SIGNIFICANCES OF DIFFERENCES BETWEEN HA AND LA

Method	HA	LA	Student's Test		Wilcoxon's Test	
			t	*p**	*T*	*p**
1st rerank (6 items)						
I ($N = 28$)	2.85	2.85				
II ($N = 28$)	2.64	2.99	3.56	<.001	140	<.01
I + II ($N = 56$)	2.75	2.92	1.75	<.05	703	<.06
2nd rerank (4 items)						
I ($N = 28$)	2.06	2.09				
II ($N = 28$)	2.03	2.20	2.17	<.03	162	<.03
I + II ($N = 56$)	2.04	2.14	1.66	<.07	712	<.08

* A one-tailed test was used.

Thus treated, the mean of the self-blame scores for the HA was 2.75, whereas that of the LA was 2.92, indicating that the HA tended to blame the self more than the LA in this test situation (Table 1). This difference proved significant at less than the .05 level of confidence.[1] However, since the tests were administered by two different methods, a breakdown by method was called for. When this was done, it was found that only in Method II was there a significant difference between the groups. For Method II, HA had a mean self-blame score of 2.64 and LA one of 2.99; this difference was significant at the .001 level of probability. There was no difference between the groups on Method I.

On the variability analysis (Table 2), the variance in ranking of each item for the four

[1] Although probability levels in the text refer to *t* tests, the data, as indicated in the tables, were also analyzed by Wilcoxon's method of unpaired replicates. The latter nonparametric test avoids the assumption of normality in the distribution of scores yet its results for this experiment are in substantial agreement with the *t* test.

TABLE 2
MEAN VARIABILITY SCORES FOR GROUPS WITH SIGNIFICANCES OF DIFFERENCES BETWEEN HA AND LA

Method	HA	LA	Student's Test t	Student's Test p*	Wilcoxon's Test T	Wilcoxon's Test p*
1st rerank (6 items)						
I (N = 28)	1.01	1.02				
II (N = 28)	0.78	1.43	2.75	< .02	150	< .02
I + II (N = 56)	0.90	1.22	1.99	< .08	688	< .08
2nd rerank (4 items)						
I (N = 28)	0.46	0.38	0.88		186	
II (N = 28)	0.38	0.64	2.05	< .06	158	< .05
I + II (N = 56)	0.42	0.51	1.06		730	

* Two tails of the distribution used.

failures was obtained for each subject. The average variance of the six items was then obtained and this gave a variability score for each subject.[2] The mean variability score for each group was then obtained and the difference between the means tested. The mean variability score of HA was .90 and for LA it was 1.22. The p value for this difference was at the .08 level of confidence, a two-tailed test being used since contrariwise to our prediction the mean variability score of the LA exceeded that of the HA. As before, the groups were broken down as to method. When this was done the difference between HA (mean variability score 0.78) and LA (1.43) on Method II proved significant at the .02 level of significance. There was no difference on Method I.

An examination of our self-blame items indicates that the ranking of two of them could be influenced by the *S*'s level of anxiety irrespective of his propensity to blame himself or others. "Nervousness or anxiety interfered with my performance," may be a simple recognition of fact for the HA *S*s, rather than an expression of self-blame. Likewise, if the HA *S* ranks highly the item, "The presence of the examiner (i.e., I work my best when I am not observed)," it may again be an indirect admission of his anxiety rather than a tendency to project blame for failure. Therefore, the data were submitted to a second reranking with these items removed and then reanalyzed. The results (Tables 1, 2) do not appreciably alter the findings. On Method II HA *S*s significantly ranked self-blame items higher than LA. Likewise on Method II, LA were more variable than HA *S*s. On Method I there are no differences between the groups.

There were no differences between the two groups on passed tests in either the self-credit or variability of rankings and since the experiment was not primarily concerned with success items and no predictions had been made concerning performance on passed tests, those data will not be presented.

Postexperimental interviews revealed that eight *S*s had some suspicion at one time or another that success on the tasks was arbitrarily determined. Because these eight persons were evenly distributed in the HA and LA categories, the differences reported above cannot be attributed to differences in this sort of insight. An analysis of the time spent by HA and LA *S*s on the various tasks shows no appreciable difference between the two categories of *S*s with respect to this variable. Consequently, the findings of this study are not attributable to differences in this variable.

DISCUSSION

Striking in the results of this experiment is the fact that whereas clear-cut, significant differences were found between HA and LA on Method II, there were no such differences on Method I. On Method II, the HA *S* blamed the self for failure to a significantly greater extent than did the LA *S*, and the HA also proved more consistent in the direction of blame assignment than did the LA. Since the sole objective difference between Method II and Method I is in the reversed order of presentation, these results suggest that the pattern of success and failure in a test situation is a crucial factor in determining a person's attitude toward his failures. Why this difference in pattern should cause a difference in attitude can only be surmised, since the experiment was not designed to give an answer to this particular question. However, before many *S*s had been tested, the examiner's observations in the test situation convinced him that Method II was more anxiety arousing than Method I.[3] This seemed to be due to the fact that failure on the Vigotsky, the initial

[2] The authors wish to express their thanks to Dr. Robert Abelson for his suggestion of a computational short cut in obtaining this measure.

[3] The selection of subjects was so arranged that the examiner was not informed as to *S*'s classification as either HA or LA.

failure in Method I, was much less stressful than failure on any other test. With the Vigotsky, the subject was allowed one attempt at solution to be completed in less than three minutes. Most *S*s made a relatively quick but incorrect grouping of the blocks. They were then given a verbal report of failure by the examiner. In contrast to this, failure on the Form Board and Kohs Block Design, which occurred early in Method II, was much more stressful, since in both cases failure involved considerable struggling on the part of *S* accompanied by immediate, personal awareness of his errors and difficulties. Therefore, in Method I, *S* on entering the test situation meets one unstressful failure followed by two successes before he again meets a failure, whereas in Method II *S* on entering the test situation meets one success followed by two stressful failures. If, now, we assume that merely entering a novel test situation involves a certain amount of anxiety for *S*, we see that Method I may permit this initial test anxiety to drop off by *S*'s becoming accustomed to the test situation before he meets his first stressful situation. But with Method II, two stressful failures occur early in the test situation, probably before the initial test anxiety has been reduced. Therefore, the two sources of anxiety, the initial test anxiety and the anxiety induced by the stressful failures, may summate to produce a maximum amount of anxiety. In Method I the total amount of anxiety induced may not be sufficient to cause an evocation of the self-relevant responses with their elements of self-blame, whereas with Method II the greater total amount of anxiety induced may be sufficient to evoke self-relevant responses. This importance of the pattern of success and failure in determining a person's attitudes towards his failures parallels the results Hutt obtained with standard and adaptive methods of administering the Stanford-Binet (2).

Since HA proved more constant in direction of blame assignment than LA, a re-examination of our assumptions is in order. It is quite possible that on Method II, the HA group experienced, contrary to our expectations, a uniformly high level of anxiety throughout the test situation. Such an analysis of the variability results, if allowed, would not alter our basic theoretical assumptions but would rather question the specific, auxiliary assumption made for our experimental design, viz., that variability in the level of anxiety and hence of the propensity for self-relevant responses on the part of HA would occur in our test situation.

REFERENCES

1. GORDON, E. M., & SARASON, S. B. The relationship between test anxiety and other anxieties, (to be published).
2. HUTT, M. L. A clinical study of "consecutive" and "adaptive" testing with the revised Stanford-Binet. *J. consult. Psychol.*, 1947, **11**, 93–103.
3. MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.
4. MANDLER, G., & SARASON, S. B. The effect of prior experience and subjective failure on the evocation of test anxiety. *J. Pers.*, 1953, **21**, 336–341.
5. SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.
6. SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810–817.
7. SARASON, S. B., & GORDON, E. M. The test anxiety questionnaire: scoring norms. *J. abnorm. soc. Psychol.*, 1953, **48**, 447–448.

Received January 21, 1954.

EINSTELLUNG EFFECT AND GOAL-SETTING BEHAVIOR[1]

NORMAN I. HARWAY

University of Rochester

Some aspects of goal-setting behavior have been presented in the literature as indicators of rigidity. Fisher (4) refers to lack of change in goals after failure as a sign of "ego level" rigidity while McAndrew (12) refers to variability in goal setting as a sign of rigidity in Lewin's sense of the term. Luchins (11) notes inconclusive results in a study of the relationship between Einstellung effect and level of aspiration. The present study investigated the latter problem: the relationship between Einstellung effect as a measure of problem-solving rigidity and goal-setting behavior.

Goldstein's (7) concept of "secondary rigidity" served here as the basis for hypothesizing the nature of the relationship between Einstellung effect and goal-setting behavior. This concept holds that an individual manifests behavioral rigidity when, unable to cope with a task, he sticks to performances which he is able to fulfill. The rigidity of behavior derives from the need to produce an adequate performance. One might argue that differences in susceptibility to Einstellung effect between two subjects faced with the same objective conditions are a function of a feeling of adequacy. The individual whose need for adequacy or success or whose fear of inadequacy or failure is great, having learned an adequate solution for the problems, tends not to "see" or seek alternative or more efficient methods of problem solution.

What should be the behavior in a level of aspiration situation of an individual who feels inadequate to the task? Two alternative behavioral patterns seem possible for such an individual. One would be not to permit himself to become involved in the situation—to state an aspiration level and stick to it, ignoring performance scores. The other would be to overreact—constantly shifting his aspiration level in an attempt to come to terms with his performance.

The hypotheses for this investigation were, then, that the rigid and nonrigid groups would exhibit different behavior in the level of aspiration situation and that the members of the rigid group would obtain either higher or lower scores on measures of variability of goal setting.

METHOD

Instruments

1. Problem-solving rigidity: water-jar series. In spite of the wide variety of measures of Einstellung effect now in use (1, 10, 14) the major instrument in this area remains the Luchins water-jar problems or an adaptation of this technique. With two exceptions, the water-jar problems used in the present investigation were those reported by Cowen (2).[2]

2. Level of aspiration. Subjects (*Ss*) were given five practice and 15 test trials on each of three tasks. The *S*'s score was reported to him after each trial and before each test trial *S* stated the score that he expected to obtain on that trial. The three tasks were:

a. Rotter Aspiration Board. The board and procedure are described by Rotter (15). The board was modified by eliminating the series of regularly spaced depressions preceding the numbered units (9).

b. Water-jar task. The *S* was presented with cards containing two water jar problems. Each card represented one trial. The score was the time taken to complete a card.[3]

c. Hidden word task. This task was designed to be somewhat similar in content to hidden-word problems often used as a measure of rigidity.[4] The procedure was similar to that of the level of aspiration water-jar task. Each card contained three problems. In a line of six letters *S* had to find a three-letter word reading from left to right.

[1] This paper is a portion of a dissertation submitted to the Department of Psychology of the University of Rochester in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his appreciation to Dr. Emory L. Cowen for the advice and guidance he supplied during the course of the study and to Dr. E. Lowell Kelly for his criticisms and suggestions as to the presentation of this paper.

[2] The specific series of test problems used is presented elsewhere (8). The series consisted of two practice problems, a control problem (solution by a long and short method), five set problems (solvable by the long method only), four critical problems (similar to the control problem), an extinction problem (solvable by a new method), and a fifth critical problem. The score was the number of critical problems solved by the short method.

[3] These problems and those of the level of aspiration hidden word task are to be found in the original report (8).

[4] An attempt was made to develop a verbal measure of Einstellung effect. This measure contained certain weaknesses which became apparent during the course of the study and is, therefore, not reported here.

Procedure

The Einstellung task was administered to a group of 118 students in introductory psychology classes at the University of Rochester. Fifty-two students who either used the long solution on the control problem, or omitted or made arithmetic errors on any except the extinction problem, were dropped from further consideration. The remaining 66 *S*s were predisposed to the shorter method of solution, completed the same number of set problems, and solved all the critical problems by the longer or shorter method.

The possible range of scores on the Einstellung series was from zero to five. Thirty-two *S*s who obtained scores of zero and one (none or one short solution) were categorized as "rigid" while 28 who obtained scores of four and five were categorized as "nonrigid." The level of aspiration tasks were administered to these 60 *S*s in individual sessions in the following order: Rotter Aspiration Board, water-jar level-of-aspiration task, hidden-word level-of-aspiration task.

Level-of-Aspiration Measures

Eleven aspects of *S*'s behavior were "scored" for each of the level-of-aspiration tasks. The last seven of the measures listed relate to the constancy or variability of *S*'s goal-setting behavior.

1. Discrepancy score: The algebraic mean of the 15 discrepancies between an *S*'s estimate for the next trial and his performance score on the immediately preceding trial.

2. Difference score: The median of *S*'s performance scores from the last practice trial through the fourteenth test trial subtracted from the mean of his 15 aspirations.

3. Mean aspiration: The mean of *S*'s estimates.

4. Number of unusual shifts: The number of times *S* raised his estimate after failure or lowered it after success.

5. Variance of discrepancy score: The mean of the squared deviations of each of the 15 discrepancies from the mean discrepancy.

6. Variance of aspiration: The mean of the squared deviations of each of the estimates from the mean estimate.

7. Number of shifts: The absolute number of times *S* changed his estimate from one trial to the next.

8. Number of aspirations: The number of *different* estimates mentioned by *S*.

9. Sum size of shifts: The sum of the points involved in the shifts from one estimate to another regardless of direction.

10. Mean size of shift: sum size of shifts divided by number of shifts.

11. Variance of size of shift: The sum of the squared deviations of each size of shift from mean size of shift divided by number of shifts.

RESULTS

Table 1 presents the variance ratios (F) between the rigid and nonrigid groups for the 11 level-of-aspiration measures. Six of the 11 measures showed significant differences for at least two of the three level-of-aspiration tasks.

TABLE 1

VARIANCE RATIO (F) AND PROBABILITY VALUE* BETWEEN RIGID (N = 32) AND NONRIGID (N = 28) GROUPS ON LEVEL-OF-ASPIRATION MEASURES

Measure	RAB†		WJ†		HW†	
	F	p	F	p	F	p
1. Discrepancy score	1.197		1.340		1.075	
2. Difference score	1.181		2.608	.02	7.422	.002
3. Mean aspiration	1.290‡		2.411	.02	10.900	.002
4. Number of unusual shifts	1.120		1.703‡		2.455‡	.02
5. Variance discrepancy score	1.038		2.272	.05	1.666	
6. Variance of aspiration	5.533	.002	5.092	.002	77.897	.002
7. Number of shifts	1.327		1.308‡		1.048‡	
8. Number of aspirations	1.158		1.633‡		1.096	
9. Sum size of shifts	5.137	.002	1.177		7.226	.002
10. Mean size of shift	3.314	.01	1.760		5.567	.002
11. Variance of size of shift	8.239	.002	6.961	.002	90.328	.002

* These are double the values presented in the tables for F. The variance ratio is considered a two-tailed test. For discussion of this point see McNemar (13, p. 230).

† RAB—Rotter Aspiration Board; WJ—water-jar task; HW—hidden-word task.

‡ Nonrigid group has the greater variance. In all other cases the rigid group has the greater variance.

In each instance the rigid group showed the greater variance. The measures distinguishing the two groups were: variance of size of shift, variance of aspiration, sum size of shifts, mean size of shift, difference score, and mean aspiration. With the exception of the last named score, the rigid group exhibited greater heterogeneity on all three tasks. Two other measures differentiated the groups with regard to variance on one of the tasks. The nonrigid group was more heterogeneous in the number of unusual shifts on the hidden word task while the rigid group seemed to be more heterogeneous with regard to the variance of discrepancy score on the water-jar task.

The conventional t ratio was not always applicable to testing the difference between group means. Sixteen of the 33 variance ratios reported indicated that in those instances the assumption of homogeneity of variance between the two groups must be rejected. In these 16 instances the test of the hypothesis that the means of the two groups were from populations having a common mean irrespective of their variances was made by an approximation method suggested by Cochran and Cox and described by Edwards (3).

TABLE 2

t RATIO AND PROBABILITY FOR THE DIFFERENCE BETWEEN MEANS OF THE RIGID (N = 32) AND NONRIGID (N = 28) GROUPS ON LEVEL-OF-ASPIRATION MEASURES

Measure	RAB		WJ		HW	
	t	*p*	*t*	*p*	*t*	*p*
1. Discrepancy score	.304	.76	.183†	.86	1.465†	.16
2. Difference score	.477†	.64	1.422*†	.17	2.684*†	.02
3. Mean aspiration	.681†	.90	2.204*†	.04	1.708*†	.10
4. Number of unusual shifts	.535†	.60	1.130†	.27	1.609*†	.11
5. Variance of discrepancy score	1.661	.11	.571*	.58	.191†	.85
6. Variance of aspiration	2.423*	.03	1.727*	.10	1.682*	.11
7. Number of shifts	.706†	.48	1.044†	.30	.304†	.76
8. Number of aspirations	.593†	.56	.269	.79	.230†	.82
9. Sum size of shifts	2.221*	.04	1.407	.17	1.655*	.11
10. Mean size of shift	3.390*	.01	2.328	.03	1.743*	.10
11. Variance of size of shift	2.339*	.03	1.630*	.12	1.366*	.19

* Assumption of homogeneity of variance is not fulfilled. Critical ratio and approximation of probability value are reported. For discussion of this point see Edwards (3, p. 168).

† Nonrigid group has larger mean. In all other cases rigid group has larger mean.

Table 2 gives *t* ratios and probability values for the differences between the means of the rigid and nonrigid groups. Only one difference was significant at the .01 level—mean size of shift on the Rotter Aspiration Board—and another six were different at less than the .05 level. Eight of the measures, however, showed the same directional trend on all three tasks.

The difference between group means was significant beyond the .05 level for one of the tasks on five of the eight measures just mentioned, and one measure showed significant differences on two of the tasks. These measures can be listed in rough order according to the apparent extent of the significance of the differences between the group means: (*a*) mean size of shift, (*b*) variance of aspiration, (*c*) sum size of shifts, (*d*) variance of size of shift, (*e*) mean aspiration, (*f*) difference score. These are the six measures for which heterogeneity of variance between groups was generally demonstrated.

Additional facts gleaned from an examination of the range of scores on the six differentiating measures were: (*a*) for all measures there was a good deal of overlap between the scores of the two groups, and (*b*) for at least three of the four measures of variability the lower end of the range of scores tended to coincide for both groups.

TABLE 3

CORRELATION COEFFICIENTS BETWEEN LEVEL OF ASPIRATION SCORES ON THREE LEVEL-OF-ASPIRATION TASKS (N = 88)

Measure	$r_{RAB\text{-}WJ}$	*p*	$r_{RAB\text{-}HW}$	*p*	$r_{WJ\text{-}HW}$	*p*
2. Difference score	.14		.24	.05	.46	.001
3. Mean aspiration	−.07*		−.16*		.36	.001
6. Variance of aspiration	.32	.01	.51	.001	.26	.02
9. Sum size of shifts	.34	.01	.47	.001	.47	.001
10. Mean size of shift	.22	.05	.50	.001	.35	.001
11. Variance of size of shift	.51	.001	.19		.23	.03

* High score on RAB means high aspiration. Low score on WJ and HW means high aspiration. These coefficients, then, imply positive relationship with regard to aspiration.

Intercorrelations of comparable scores on the three level-of-aspiration tasks for the six measures differentiating the two groups are given in Table 3. The correlation coefficients were based on 88 *Ss*.[5]

DISCUSSION

The difference between the nonrigid and rigid groups both with regard to variances and means on six of the 11 level-of-aspiration measures suggested that the two groups were drawn from different populations with respect to attributes tapped by these aspects of goal-setting behavior. It was apparent that the greater heterogeneity of scores in the rigid group was a function of a number of *Ss* producing extreme scores in one direction. The differentiations to be made between the nonrigid and rigid groups are then really between the nonrigid group and the more extreme members of the rigid group.

The level-of-aspiration measures on which the scores of the two groups of this study differed may be separated into two categories: measures relating to variability of behavior and measures relating to height of aspiration. The individuals of the rigid group were more variable in behavior on the level-of-aspiration tasks: (*a*) when they changed their estimate from one trial to the next, they made larger shifts; (*b*) they tended to vary more about

[5] The 88 *Ss* included the 60 chosen by the water-jar Einstellung series and 28 who were chosen by a hidden-word Einstellung series.

their individual mean estimate, that is, they varied their estimates over a wider range; (*c*) if one ignores the number of times *S*s shifted their goals, the total number of units involved in all the shifts made by *S*s was greater for the members of the rigid group; (*d*) in shifting their goals the members of the rigid group tended to utilize shifts of many more different sizes than did the members of the nonrigid group.

Rigidity in problem solving thus seems to be related to greater variability of goal-setting behavior. Of the two level-of-aspiration measures of rigidity that are referred to in the literature (5, 6, 9, 15) one was negatively related to Einstellung effect—the rigid group had the greater sum size of shifts—while the other—number of shifts—did not significantly differentiate the two groups.

Members of the rigid group did exceed the upper end of the range of scores of the nonrigid group. If a feeling of inadequacy was indeed what underlay the behavior of at least certain of the members of the rigid group, the question should be raised as to why some members of this group did not show the other extreme of goal-setting behavior, i.e., state an aspiration level and not change it regardless of the performance scores. In the formulation of the hypothesis this was suggested as an alternative indicator of an *S*'s feeling inadequate to the task. Also, while the lower aspiration level in the rigid group is consistent with feeling inadequate, an *S* might try to compensate for such feelings by setting a goal much higher than his performance. A possible explanation for the failure to find such behaviors is that they may be a function of greater degrees of rigidity and stronger doubts as to adequacy than those touched upon in the present study.

SUMMARY

Goal-setting behavior of rigid *S*s was compared to that of nonrigid *S*s. Rigidity was measured by performance on the water-jar series; goal-setting behavior was evaluated on three level-of-aspiration tasks: Rotter Aspiration Board, a water-jar task, and a hidden-word task.

Hypotheses regarding the level-of-aspiration behavior of the rigid and nonrigid groups of *S*s were deduced from Goldstein's concept of secondary rigidity. The two groups were found to differ on six of the 11 measures derived from the level-of-aspiration situation. Four measured variability of goal-setting behavior and two described height of aspiration. The rigid group was more heterogeneous on all six measures; as a group they scored higher on the measures of variability and set lower aspiration levels.

Conclusions were limited by the great overlap of scores in the two groups. It was suggested that a feeling of inadequacy for the tasks involved could explain the level-of-aspiration performance of some of the members of the rigid group.

REFERENCES

1. COWEN, E. L. A study of the influence of varying degrees of psychological stress on problem-solving rigidity. Unpublished doctor's dissertation, Syracuse Univer., 1950.
2. COWEN, E. L. The influence of varying degrees of psychological stress on problem solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512–519.
3. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
4. FISHER, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
5. FRANK, J. D. Individual differences in certain aspects of the level of aspiration. *Amer. J. Psychol.*, 1935, **47**, 119–128.
6. FRANK, J. D. Recent studies of level of aspiration. *Psychol. Bull.*, 1941, **38**, 218–226.
7. GOLDSTEIN, K. Concerning rigidity. *Charact. & Pers.*, 1943, **11**, 209–226.
8. HARWAY, N. I. Personality variables in problem solving rigidity inferred from behavior in the level of aspiration situation. Unpublished doctor's dissertation, Univer. of Rochester, 1952.
9. KLUGMAN, S. F. Emotional stability and level of aspiration. *J. gen. Psychol.*, 1948, **38**, 101–118.
10. LUCHINS, A. S. *An examination for rigidity of behavior*. New York: Veterans Administration, 1948.
11. Luchins, A. S. Rigidity and ethnocentrism: a critique. *J. Pers.*, 1949, **17**, 449–466.
12. MCANDREW, HELTON. Rigidity and isolation: a study of the deaf and blind. *J. abnorm. soc. Psychol.*, 1948, **43**, 476–494.
13. MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
14. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. Unpublished doctor's dissertation, Univer. of California, 1947.
15. ROTTER, J. B. Level of aspiration as a method of studying personality: II. Development and evolution of a controlled method. *J. exp. Psychol.*, 1942, **31**, 410–429.

Received February 2, 1954.

PROBLEM-SOLVING RIGIDITY AND DECISION TIME[1, 2]

DAVID LEVINE

New York University

A WIDE variety of psychological research indicates that a number of people have difficulty adjusting to new situations because of their tendency to cling to well-established but no longer appropriate patterns of behavior. This phenomenon is generally referred to as *behavioral rigidity*. Some writers have suggested that rigidity is largely a function of excessive caution related to the fear of making new responses, while others indicate that the tendency to respond quickly, by limiting the variety of available responses, leads to rigidity (1, 4). Thus, two apparently contradictory explanations have been offered. It is possible that both explanations are valid, i.e., in some people excessive caution may lead to rigidity, in others it may stem from impulsivity. The present study, employing the familiar Einstellung water-jar problems as the measure of rigidity and a simple discrimination situation as the measure of impulsivity-caution, is an investigation of these various hypotheses.

PROCEDURE

The subjects (*Ss*) were 100 male veterans, between the ages of 19 and 45, who volunteered to take part in the study. All of the *Ss* had a medical disability related to their service experience, but none of the *Ss* was considered to be psychotic or grossly impaired organically. The impulsivity-caution, or decision-time, test was administered first. In this test the *Ss*, who were seen individually, were shown 22 cards, on each of which two figures had been drawn. The *S* was required to say which of the two figures was larger. On 12 of the 22 cards the figures were objectively equal. The individual's response time to all cards was recorded. His decision time score was his average response time to the ambiguous cards (conflict time) less his average response time to the nonambiguous cards (base time). This procedure, suggested by Festinger and Wapner (2), was followed in order to control for interindividual variability in reaction time. The odd-even reliability coefficient for this score was .52; this is similar to that obtained by Festinger and Wapner.

The *S* was next given the water-jar problems. The series of problems used in the present study consisted of five set-producing problems soluble only in an indirect manner followed by a single critical problem which could be solved only in a simple and direct manner. The time spent on each problem was recorded. The rigidity score was the ratio of the time taken to solve the final critical problem to the average time on the set-producing problems. This procedure furnishes a measure of control for interindividual differences in arithmetical ability.

RESULTS

In order to investigate the data, each *S* was put into one of three groups depending on his decision-time score. The 25 *Ss* with the longest decision-time scores (ranging from 9.4 to 39 seconds) formed the Cautious group; the 25 *Ss* with the shortest decision-time scores (ranging from .8 to 2.4 seconds) formed the Impulsive group; the 50 *Ss* whose scores ranged from 2.4 to 9.4 seconds formed the Moderate group. The total sample was also divided according to whether the disability was psychiatric or nonpsychiatric. Since the rigidity scores yielded a markedly skewed distribution, they were submitted to a logarithmic transformation prior to statistical analysis. The rigidity scores for the subgroups are shown in Table 1. The data were submitted to an analysis of variance; none of the differences between groups was found to be significant at the .10 level.

CONCLUSION

The present study failed to reveal any significant relationship between rigidity and

TABLE 1

RIGIDITY SCORES

Group	Impulsive	Moderate	Cautious	Total
Nonpsychiatric	.77 (13)*	.51 (34)	.61 (15)	.60 (62)
Psychiatric	.54 (12)	.70 (16)	.68 (10)	.67 (38)
Total	.67 (25)	.57 (50)	.70 (25)	.63 (100)

* Numbers in parentheses refer to the *N* upon which each mean is based.

[1] Based on a doctoral dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University, 1953. The writer wishes to acknowledge his gratitude to Dr. Irwin N. Katz who served as chairman of the Thesis Committee.

[2] From VA Hospital, Lyons, New Jersey.

decision time. It is possible that the measures employed did not tap the variables under consideration. Using a simple discrimination situation in order to measure a complex variable like impulsivity-caution involves many assumptions. The heterogeneous nature of the sample employed may have also served to conceal the hypothesized relationships. Nevertheless, the present results suggest that attempts to explain behavioral rigidity may have to resort to more complex concepts than impulsivity and caution.

REFERENCES

1. CATTELL, R. B., & WINDER, A. E. Structural rigidity in relation to learning theory and clinical psychology. *Psychol. Rev.*, 1952, **59**, 23–40.
2. FESTINGER, L., & WAPNER, S. *A test of decision time: reliability and "generality."* Washington, D. C.: U. S. Dept. Commerce, 1947. Civil Aeronautics Administration. (CAA Div. Res. Rep. No. 48, 1945; Publ. Bd. No. 50320.)
3. LEVINE, D. Problem-solving rigidity and impulsivity-caution. Unpublished doctor's dissertation, New York Univer., 1953.
4. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.

Received February 4, 1954.

RIGIDITY AND REACTIVE INHIBITION[1]

ROBERT B. MALMO AND HARVEY WALLERSTEIN[2]

Allan Memorial Institute of Psychiatry and McGill University

RIGIDITY means that a reaction pattern is repeated more than it should be, according to some criterion of what is normal in that situation. The repetitive response is so strong that other alternative responses cannot compete successfully with it, and so they remain below reaction threshold. If the repetitive response could be inhibited sufficiently, then, as a matter of course, some stronger competing response would appear, and the behavior would be less repetitive and less rigid.

There are at least two kinds of inhibition, interference and adaptation (4) or reactive inhibition (5). A recent example of the interference approach is found in the work of Ferguson and his students (11, 13) who hypothesized that rigid persistence in behavior may be produced by prior learning which interferes with evocation of the more appropriate reactions. Luchins' hypotheses concerning Einstellung also stress the concept of interference (8).

Reactive inhibition, the decrement in a response which is a consequence of its repeated elicitation, has been neglected in investigations of rigidity. As far as we know, the present experiment is the first one to be designed specifically to study the relationship between rigidity and reactive inhibition. We selected a simple motor reaction with which we were familiar from previous studies: button pressure in response to painful stimulation. Reactive inhibition of the button-pressure response will appear as a progressive lessening of the reaction with its repetition (becoming slower, shorter, and weaker). The next thing which is required in this experimental design is to compare a group of rigid individuals with a group of controls in order to test the hypothesis that rigid persons are less affected by reactive inhibition than normally flexible persons. On clinical grounds[3] (6, 7, 14) we assumed that psychoneurotic patients as a group were more rigid than controls.

A more concrete statement of our hypothesis is that psychoneurotics will show less decrement than normals in the strength of the button pressure response, as a function of its repeated elicitation.

METHOD

Subjects. The patient group comprised 29 patients (11 males and 18 females) from the Institute wards. Only six of the 29 patients were anxiety patients, a category which we have previously defined (9, p. 9ff). Hysterical components were the most prominent features in fifteen, depressive features in seven, and schizoid features in one. It is perhaps worth stressing that this is not primarily an "anxious" group, because, for this and other reasons, we do not favor the view that results to be described can be ascribed to "anxiety." Nineteen control subjects (*Ss*) (14 males and 5 females) were employed. Thirteen *Ss*, eight men and five women were university students. The remaining six male controls were Royal Canadian Air Force personnel. Patients ranged in age from 18 to 56 years (median of 31); controls from 19 to 36 (median age of 21 years). Unfortunately, it was not possible to obtain a better matching between groups for age and sex. This was largely due to the way the *Ss* were selected from a larger experimental population, and the other criteria for selection, which are described below, were deemed more important than matching precisely for age and sex. However, all conclusions from present data must be examined with these differences in mind, and their influence must, in all possible instances, be estimated. It will be noted, for example, that the patient group is an older group than the controls. As will be clear later, when the results are presented, age did not appear to be a significant factor in the present study.

Apparatus and procedure. Procedure in the present experiment was similar to the pain stimulation method previously described (9). Light from a 500-watt projection lamp was focused by means of a condensing lens on the forehead, which had been blacked with

[1] This investigation was supported by the Research and Development Division, Office of The Surgeon General, Department of the U. S. Army, under Contract No. DA-49-007-MD-70.

[2] The authors are grateful to Drs. Alexander Bartoshuk, George Ferguson, Abraham Luchins, and Blossom Wigdor for their critical reading of the manuscript.

[3] For example, Kubie in his paper on "the repetitive core of neurosis" argues that psychoneurosis is "pathological distortion of normal repetitiveness" (6). He is clear on the point that such repetitiveness is a characteristic feature of neurotic individuals (that they deviate in this way from "normal"), and that this is not confined to a small group (e.g., obsessive-compulsives who may, perhaps, be referred to as the extreme of extremes).

India ink. The lamp was on continuously, at a steady power consumption of 340 watts (slightly above the pricking pain threshold of Hardy, Wolff, and Goodell [3]). Ten 3-second stimulations were presented to the forehead at 1-minute intervals. The 1-minute interval was chosen because Hardy, Wolff, and Goodell reported that ten exposures to a threshold pain stimulus 1 minute apart had no effect upon pain threshold (3, p. 109). They found that repeated stimulation of the same area of skin, if done in rapid succession, caused lowering of the pain threshold and local hyperalgesia; they attributed the change in pain threshold under repetitive stimulation to a raised skin temperature and to the hyperalgesia. This control study of Hardy, Wolff, and Goodell also provides evidence that the conditions of our experiment (which were practically identical with theirs) were not such as to produce the kind of receptor change which would distort our reaction measurements. In particular, adaptation of pain receptors does not appear under the conditions of their experiment. Since there is good evidence that pain thresholds (of perception) in psychoneurotics are not significantly different from normals (2), we may apply these findings of Hardy, Wolff, and Goodell to our psychoneurotic *Ss*.

The right forefinger rested on a button which *S* was instructed to press when he felt the heat was about to become painful. It was explained that pressure on the button would not terminate the stimulus. Further instructions were to keep eyes closed, and not to withdraw from stimulation. Finger movement on the button was recorded with a Rochelle salt crystal pickup, and muscle potentials were recorded from right forearm extensors, with standard electrodes (1). Other details of apparatus and procedure are described elsewhere (10), as well as other procedures (not relevant to the present paper) which were carried out on our *Ss* in the same experimental session.

We mentioned earlier that the *Ss* in this experiment were selected from a larger group. The *Ss* in the present experiment were those who chose to press the button on every stimulation of the forehead. It may be noted that under the conditions of the present experiment, frequency of pressure on the button was not greater than the controls' in the total group of patients, as might have been expected. Actually, the reverse was true. Only 35 per cent of anxiety patients were in the group of "pressers," whereas 58 per cent of hysterics, 50 per cent of depressives, and 59 per cent of normals pressed every time. It should be noted here that these findings are the reverse of those reported earlier (9, p. 16). However, in the earlier experiment instructions appeared to lead most patients to believe that pressure on the button would terminate stimulation. In the present experiment all *Ss* were explicitly told that such was definitely not the case, that pressure would have no effect whatever on stimulation. This change in instructions appeared to alter the significance of the act, especially for "anxious" patients. It thus appeared that pressure on the button did not necessarily signify fear or apprehension. Further evidence opposed to the interpretation of button pressure as an "anxiety reaction" is the fact that controls pressed more frequently than patients. The over-all percentages (for all *Ss* in each group) were 87 per cent pressures for the control group, and 66 per cent for the patient group (chi square of 12.26, $p < .001$). Actually, whatever pain was felt was very mild (in contrast to some of the stimulations in the previous study) and extremely brief. There were 60 patients in the original group, and 32 controls.

Treatment of data. The button recorder (on which the *Ss* pressed) yielded two measures of response strength, latency, and duration of pressure. In addition to these two measures of response strength, muscle potentials from the right arm provided an ancillary measure, that of "tension" throughout the series of stimulations. Measurements were obtained from integrators of the meter type. Each meter recording represented integration of EMG's for a 10-second period. Readings were made for alternate 10-second periods throughout the test (three readings per minute). Integrator and stimulator were synchronized so that a 10-second period of integration was always initiated with the onset of each stimulus. Meter readings were converted to microvolt values by reference to calibration standards.

RESULTS

Figure 1 shows that mean latency increased progressively throughout the ten trials for the control group. This was not true of the group of patients who showed no significant change in this measure during the course of stimulations. In patients the change was from 1.32 seconds on stimulus 1 to 1.52 seconds on stimulus 10 (t of 0.98, p of 0.30). Controls' latency increased significantly from 0.99 seconds to 2.12 seconds (t of 7.90, p less than 0.001). The patient-control difference in amount of change was significant at the 1 per cent level (t of 3.39).

Figure 2 shows that mean duration of pressure on the button decreased progressively in the control group, and again, the patients pre-

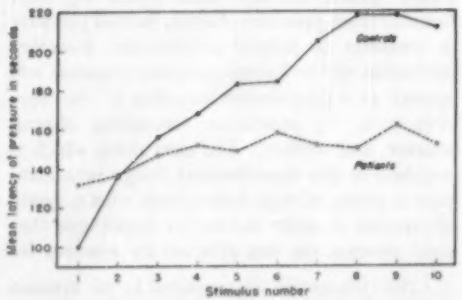

FIG. 1. COMPARISON OF PATIENTS AND CONTROLS IN MEAN LATENCY OF PRESSURE ON BUTTON IN RESPONSE TO MILDLY PAINFUL THERMAL RADIATION OF FOREHEAD.

Note that controls' latencies increased progressively throughout the series of ten stimulations, while patients' failed to show marked change.

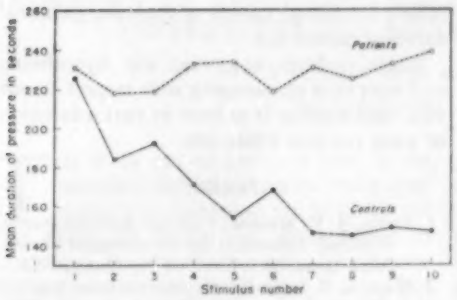

FIG. 2. COMPARISON OF PATIENTS AND CONTROLS IN MEAN DURATION OF PRESSURE ON BUTTON IN RESPONSE TO MILDLY PAINFUL THERMAL RADIATION OF FOREHEAD.

Note that duration of pressure in controls decreased progressively throughout the series of ten stimulations, while patients failed to show this downward trend.

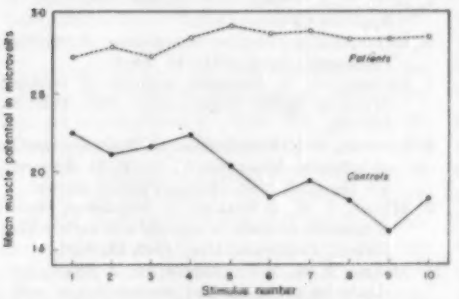

FIG. 3. COMPARISON OF PATIENTS AND CONTROLS IN MEAN RIGHT FOREARM MUSCLE "TENSION" DURING THE TEN-MINUTE PERIOD WHEN THERMAL STIMULATIONS WERE BEING PRESENTED.

Note the falling "tension" in controls, for this period, and contrast this with the "static tension" of patients.

sented a quite different picture. Patients' average duration of the first stimulation was 2.31 seconds, and it was even a little higher for the tenth stimulus, 2.39 seconds, though the increase was not significant. Controls, on the other hand, showed a steady *decrease* in duration from 2.26 seconds on stimulus 1 to 1.47 seconds on the last stimulus. This change was highly significant (t of 3.32, p less than 0.01). The patient-control difference in change was significant at the 5 per cent level (t of 2.07).

The muscle potential data are presented in Fig. 3. Though differences between patients and controls were not statistically significant, it is of interest to note that, once again, the trend in controls was steadily downward, while patients' "tension" remained unchanged thoughout the series of stimulations.

Considerations of age and sex. No significant correlations between age and the three measures of reaction change were found. For the control group, r's for age and change (from first to last stimulus) in duration, latency, and "tension" were as follows: .16, −.40, and −.04. For the group of patients the corresponding values were .27, −.05, and .17. In addition, the Olmstead-Tukey (12) nonparametric test of association which was applied to the duration and latency data failed to establish any significant relationship between either of these two variables and age.

None of the differences between patients and controls appeared attributable to the sex factor. When the data were fractionated for sex, greater change in controls than in patients appeared in the fractionated data (male controls vs. male patients, female controls vs. female patients).

DISCUSSION

Present data confirm our hypothesis that psychoneurotics show less decrement in the strength of a voluntary response as a function of its repeated elicitation. This in turn is encouraging with respect to our view that rigidity is in part, at least, a function of weak reactive inhibition. If the relationship between rigidity and reactive inhibition, which we have found under these specific experimental conditions holds generally, various other hypotheses follow. These may be put to direct test. For example, we should expect to find slower reflex habituation and slower extinction of conditioned responses in psychoneurotics than in less rigid individuals.

As a second line of attack on this problem, it may prove fruitful to resume the approach of factor analysis, with emphasis on reactive inhibition in the test construction. Here a clinical scale of rigidity should be very useful.

Finally, we may consider interpretation of our findings in terms of the "anxiety" factor. Since this factor could not be equated in the two groups of *S*s, it is important to consider how much the presumably greater anxiety in the patients may have affected the results. The main question is, how much were our results affected by specific fear of pain as heightened by generalized anxiety in the psychoneurotic

group (9)? In a previous study in which instructions were worded so that *Ss* believed that pressure on the button terminated stimulation, there was a definite relationship between degree of anxiety and frequency of button pressure (9, p. 16). However, in the present study, with instructions worded in an unambiguous way to the effect that button pressure would not turn off stimulation, quite different results were obtained. Percentage of anxiety patients pressing every time was actually much smaller than the percentages for the other group of patients and for normal controls. Furthermore, frequency of pressure was significantly higher for all controls than for all patients. This evidence is opposed to the conclusion that strength of the button pressure response is a positive and simple function of anxiety. We do not consider, therefore, that our patients in this study showed a heightened fear reaction to pain, which could explain their more rigid performance.

SUMMARY AND CONCLUSIONS

Hypothesis: Psychoneurotics will show less decrement in the strength of a voluntary response, as a function of its repeated elicitation, than normals. The two chief response measures were latency and duration of forefinger pressure on a button in response to mildly painful thermal stimulation of the forehead. A modified Hardy-Wolff-Goodell apparatus provided ten successive stimuli of 3 seconds' duration, spaced at 1-minute intervals. Muscle potentials from the responding arm provided a third related measure.

Subjects were 29 psychiatric patients and 19 controls.

In all three measures, controls showed progressive weakening of reaction (increasing latencies, decreasing durations, and decreasing muscle "tension"). Patients, on the other hand, failed to show these changes in reaction to repeated stimulation. In contrast to the falling (or rising) curves of controls, patients' curves appeared flat.

These findings supported the hypothesis, and were thus encouraging with respect to our view that rigidity is at least in part a function of weak reactive inhibition.

REFERENCES

1. DAVIS, J. F. *Manual of surface electromyography.* Montreal: Laboratory for Psychological Studies, Allan Memorial Institute of Psychiatry, 1952.
2. HALL, K. R. L. Studies of cutaneous pain: A survey of research since 1940. *Brit. J. Psychol.*, 1953, **44**, 279–294.
3. HARDY, J. D., WOLFF, H. G., & GOODELL, H. *Pain sensations and reactions.* Baltimore: Williams & Wilkins, 1952.
4. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
5. HULL, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
6. KUBIE, L. S. The repetitive core of neurosis. *Psychoanal. Quart.*, 1941, **10**, 23–43.
7. LUCHINS, A. S. Proposed methods of studying degrees of rigidity in behavior. *J. Pers.*, 1947, **15**, 242–246.
8. LUCHINS, A. S. Examination for flexibility-rigidity of behavior. Montrose, N. Y.: F. D. Roosevelt VA Hospital, 1950. (Mimeographed manual.)
9. MALMO, R. B., & SHAGASS, C. Physiologic studies of reaction to stress in anxiety and early schizophrenia. *Psychosom. Med.*, 1949, **11**, 9–24.
10. MALMO, R. B., WALLERSTEIN, H., & SHAGASS, C. Headache proneness and mechanisms of motor conflict in psychiatric patients. *J. Pers.*, 1953, **22**, 163–187.
11. OLIVER, J. A., & FERGUSON, G. A. A factorial study of tests of rigidity. *Canad. J. Psychol.*, 1951, **5**, 49–59.
12. OLMSTEAD, P. S., & TUKEY, J. W. A corner test for association. *Ann. math. Statist.*, 1947, **18**, 495–513.
13. SCHEIER, I. H., & FERGUSON, G. A. Further factorial studies of tests of rigidity. *Canad. J. Psychol.*, 1952, **6**, 18–30.
14. WIGDOR, BLOSSOM, T. An approach to personality through an investigation of perceptual behavior. Unpublished doctor's dissertation, McGill Univer., 1952.

Received February 10, 1954.

RESPONSES TO REPEATED FRUSTRATIONS OF YOUNG CHILDREN DIFFERENTIATED ACCORDING TO NEED AREA

NANCY BARKER OTIS AND BOYD McCANDLESS

Iowa Child Welfare Research Station

THIS study was suggested in part by the theoretical formulations concerning frustration and aggression set forth by the Yale group (2). Frustration was assumed to occur when there had been some interference that prevented an organism from making the necessary instrumental acts to complete a goal response, and is conceived to result in "a special secondary frustration-induced instigation, the goal response to which is aggression" (6, p. 343). It was also considered possible for frustration, in terms of previous learning of the organism, to instigate other behaviors. These, however, were not necessarily considered to be frustration reducing. Miller (6, p. 339) states: "If . . . the first responses do not lead to a reduction in the original instigation, then the instigations to them will tend to become weakened through extinction so that the next most dominant responses, which may or may not be aggression, will tend to occur. From this analysis it follows that the more successive responses of non-aggression are extinguished by continued frustration, the greater is the probability that the instigation to aggression eventually will become dominant so that some response of aggression actually will occur."

Hence it has been predicted here that other than aggressive responses would decrease in strength and/or number during prolonged or repeated frustration, whereas dominant-aggressive responses, being frustration reducing, would increase in number and/or strength.

The concern here is with two classes of behaviors: dominant-aggressive (Ag) on the one hand and complaisant-submissive (Su) on the other hand. The most general hypothesis follows the Yale thinking: under conditions of frustration, the Ag behaviors (in this case a score based on number and judged intensity) will increase, the Su behaviors (also a score) will decrease.

In other words, the authors' intention is not so much to describe the behaviors occurring because of or concomitant with frustration as defined by a task, but to investigate the problem of the conditions under which one set of responses to frustration rather than another can be predicted to occur and vary during frustration. Hence, additionally, this study concerned itself with an attempt to classify children in terms of the strengths of two different social needs, on the assumption that these provide different motivational systems for reaction to frustration, hence will lead to different frustration behavior. The needs investigated here were: power-dominance (PD/n) and love-affection (LA/n).

Generally speaking, the following hypotheses, which are stated more precisely in the Results section, were formulated: (*a*) Ag responses (score) will increase during frustration; (*b*) Su responses (score) will decrease during frustration. At a more clinical level, and going beyond the Yale hypothesis, (*c*) children rated high in PD/n will show more Ag and less Su behavior in a frustration series than will children ranked low, and (*d*) they will also increase more in Ag behavior during the series than will children ranked low. On the other hand, (*e*) LA/n ratings will be positively related to Su scores and negatively related to Ag scores during a frustration sequence taken as a whole.

METHOD

Experimental procedures. The frustration task selected consisted of two "blocking techniques" devised by Lerner (4). Essentially these consist of having the experimenter (*E*) start from one end of a road constructed of preschool blocks, the subject (*S*) from the other, with the issue of *E* meeting *S* being the area of conflict or frustration of *S* (upon meeting, *E* asks, "How shall my car [or doll] pass?"). Eight repetitions of the situation, the first and last three being "car meets car" and the fourth and fifth being "doll meets car" provided the standard experimental situation for each child. This situation might be subjectively characterized as being one of standardized, mild, repeated frustration, but one in which the children did not appear to be more than minimally involved. The required time per child was approximately six minutes, with a four- to ten-minute range.

A pretested manual of scoring for frequency and intensity of Ag and Su scoring was used for a total of five different observers (*Os*) for the three experimental groups, A, B, and C.

TABLE 1

AGGRESSION-DOMINANCE (AG) AND SUBMISSION-COMPLAISANT (SU) SCORE CHANGES FROM THE FIRST TO SECOND HALF OF A FRUSTRATION SERIES

Measure	Group		
	A	B	C
Ag			
Mean change	2.4	2.7	1.8
SD change	4.264	3.667	5.870
t	2.212*	2.661*	1.715
Combined *p*		<.01	
Su			
Mean change	−0.7	−1.6	−1.4
SD change	1.902	2.459	3.073
t	1.518	2.416*	6.606**
Combined *p*		<.01	

Note.—Combined *p*'s are computed according to Lindquist (5, pp. 148–156), *t*'s according to Edwards (3).

* $p < .05$.

** $p < .01$.

For Groups A and B, PD/n and LA/n were rated by the paired-comparison method by teachers familiar with the children but naive as far as the experiment was concerned. One pair of teachers for each group independently rated PD/n, another pair LA/n. These two variables were also judged by two pairs of clinical psychologists from CAT (1) protocols, each pair working with one of the two variables for both Groups A and B. They operated "blind" except for information about the child's group and sex. Details of these methods may be obtained from the second author.

Reliabilities of observers (*Os*) for Groups A and B for total Ag and total Su scores showed Pearson *r*'s from .94 to .97. On the assumption that this extremely high observer agreement for the three *Os* concerned with these groups might be due to sophistication about the experimental design, a third group (C) was run with new and naive *Os* (as to this experiment). Their reliabilities for total Ag were .84 and for total Su .91 (Pearson *r*'s).

The reliabilities of the teacher ratings for PD/n for Groups A and B respectively were .86 and .84; and for LA/n .60 and .63, uncorrected for attenuation. The reliabilities of the clinician ratings for PD/n inferred from the CAT were .44 and .31 for groups (all Pearson *r*'s). Since these last reliabilities did not differ from zero, "agreed upon pairs" (i.e., children who the raters agreed belonged above or below the mean in need strength) were used in the analysis. There were seven high and seven low, seven high and seven low, for LA/n for Groups A and B respectively. The same figures for PD/n were nine and nine; and nine and eight.

Subjects. The major portion of the present study was conducted using two groups of children (Group A the elder, Group B the younger) who were enrolled in the Laboratory Preschools of the Iowa Child Welfare Research Station. A majority of the *Ss* were children of students or faculty at the University, and thus fall higher than average on most scales of socioeconomic-intellectual status. All of these children with one exception had been in attendance at the Preschools since the fall preceding the experiment, or approximately eight months. Many had previously attended one full year and the experimental year was the third year for some. A subsection of the study included Group C, made up of children all new to the Preschools with one exception (a child not included in the original experiment, since he was not enrolled until after its completion). These children, with this exception, were all without prior preschool experience and had been in attendance only a week to 10 days when they were placed in the frustration situation. Most were temporary, summer school children, whose parents were attending summer school, although some were regular Iowa City residents.

A total of 63 *Ss* was used, 31 boys and 32 girls, approximately evenly distributed by sex and number in Groups A, B, and C. The average ages of the groups were 62, 52, and 54 months, respectively. Total age range was from 41 to 65 months.

RESULTS

Hypothesis 1 predicts that the second four trials of the eight-trial frustration sequence will result in higher Ag than will the first four trials. Table 1 summarizes the relevant data and demonstrates that Groups A and B showed changes significantly (at or below the .05 level) supporting the hypothesis, that Group C's changes were in the direction of the hypothesis at the .10 level, and that in terms of pooled probability the hypothesis was supported at less than the .01 level.

Hypothesis 2 predicts that for total groups Su will be lower for the second four than for the first four trials of a frustration series. Table 1 summarizes the data with reference to this hypothesis. Groups B and C show statistically significant shifts downward, Group A a shift (nonsignificant) in the predicted direction. The pooled probability for all three groups is less than .01.

Parenthetically, as can be noted from Table 2, while there are no significant differences in Ag or Su between Groups A and B (those children with rather prolonged attendance in a permissive preschool atmosphere), Group C, children new to preschool, showed significantly less Ag and significantly more Su than either Groups A or B.

To test hypotheses 3 to 5 it was necessary that independently rating judges (teachers familiar with the children) define in a reliable fashion children who are high in PD/n and in LA/n.

As mentioned earlier, PD/n ratings proved more reliable in terms of interrater agreement than did LA/n, perhaps because the behavior on which the ratings were based is more likely to call itself strongly to the attention of teachers responsible for the smooth conduct of a group, while LA/n manifestations may be relatively solitary and differential in nature (i.e., directed at one person, and selectively for a favorite teacher).

However, the hope that reliable ratings of PD/n and LA/n could be made from CAT protocols did not work out (interrater reliabilities, expressed as correlation coefficients, ranged from .27 to .44).

A related hope that teacher and CAT protocol rating definitions of PD/n and LA/n would be positively related (e.g., collapsible) was not fulfilled, owing to the unreliability of the CAT-based ratings. The CAT ratings did not enter into either any significant or consistent empirical relationships with any of the other data. Consequently no further CAT data are reported here.[1]

It was predicted in hypothesis 3 that children rated high for PD/n (teacher ratings) would show more Ag and less Su during the entire frustration sequence. Table 3 summarizes the data collected with reference to this hypothesis. It will be recalled that Group C undertook only the frustration experiment, and that ratings are not available for them.

Group A provided results in support of the hypothesis at a statistically significant level; Group B's results were nonsignificant but in the direction of the hypothesis, and the pooled r's demonstrated statistically significant support of the hypothesis.

Hypothesis 4 predicted greater increases in Ag from the first to the second four trials of the frustration task for children ranked high in PD/n than for children ranked low. The data relevant to this hypothesis are also summarized in Table 2 (no hypothesis was set forth about Su change as related to teacher definitions of PD/n). It can be noted that the hypothesis is supported at a statistically significant level for both Groups A and B, and that the combined r for PD/n and increase in Ag is .45 (significant at less than the .01 level). While

[1] They may be secured on request from either author.

TABLE 2

t's FOR SIGNIFICANCE OF DIFFERENCES BETWEEN GROUP MEANS FOR TOTAL AGGRESSION (Ag) AND TOTAL SUBMISSION (Su)

Groups Compared	Ag	Su
A vs. B	.107	1.101
A vs. C	2.244*	4.667**
B vs. C	2.415*	3.133**

* $p < .05$.
** $p < .01$.

TABLE 3

RELATIONS (r's) BETWEEN TEACHER RATINGS OF POWER-DOMINANCE NEEDS (PD/n) AND FRUSTRATION BEHAVIOR

Group	PD/n vs. Total Ag	PD/n vs. Ag Change	PD/n vs. Total Su	PD/n vs. Su Change
A	.68**	.48*	−.52*	−.35
B	.24	.43*	−.11	−.24
Combined	.49**	.45**	−.33*	−.30

Note.—Combined r's were computed according to the method suggested by Snedecor (7, pp. 148–156). These r's are Pearsonian.
* $p = .05$.
** $p = .01$.

no relationship was postulated between PD/n and Su change, the relationships are actually consistently negative. The pooled r (−.30) just misses significance at the .05 level.

It was proposed in hypothesis 5 that children whose teachers defined them as high in LA/n will show more total Su during frustration, but less total Ag; a further, *post hoc* check at a frankly speculative level was based on the hunch that they will be less likely to "extinguish" Su from the first to the second four trials of a frustration task, when compared with children defined as low in LA/n. Table 4 summarizes the relevant data.

Hypothesis 5 was supported at a statistically significant level for Group A and for the total group as indicated by combined r's (i.e., high LA/n goes along with more Su and less Ag behavior). Results for Group B were in the predicted direction, but statistically nonsignificant. It was also found that LA/n was negatively related to Ag change during frustration, both for Group A and for the groups combined, with Group B producing results in the predicted direction, but at a statistically insignificant level. No significant

TABLE 4

RELATIONS (r's) BETWEEN TEACHER RATINGS OF LOVE-AFFECTION NEEDS (LA/n) AND FRUSTRATION BEHAVIOR

Measure	LA/n vs. Total Ag	LA/n vs. Ag Change	LA/n vs. Total Su	LA/n vs. Su Change
Group A				
r	−.55**	−.48*	.50*	.22
p	<.01	<.05 >.01	<.05 >.01	NS
Group B				
r	−.06	−.26	.14	.19
p	NS	NS	NS	NS
Combined				
r	−.33*	−.38*	.33*	.20
p	<.05 >.01	<.02 >.01	<.05 >.01	NS

Note.—Combined r's were computed according to the method suggested by Snedecor (7, pp. 46-47). These r's are Pearsonian.

* $p < .05$.

** $p < .01$.

relationship with Su change was demonstrated by either group taken alone or in combination, although the r's are positive.

Finally, as a matter of normative interest, sex differences were computed for Ag and Su changes. There were no significant ones (boys showed Ag changes of 2.42 score points, girls 2.27; boys showed Su changes of −1.42 score points, girls of −1.04 score points). However, there were significant sex differences in the expected (usual) direction in terms of total Ag and Su behavior shown in the frustration experiment.

DISCUSSION

First, it seems to the authors that this study provides a certain testimonial to the power of theory, particularly in the case of the first two hypotheses. By the use of the Yale theories, which are by no means recent, remarkably accurate predictions of the course of Ag behavior and Su behavior during repeated but mild frustration were provided. Ag behavior scores, which may be thought to represent behavior considered by the theory to be reducing for frustration, did indeed increase as the frustration (defined by repetition of the frustration situation) became stronger.

But Su behavior scores, representing behavior considered by the theory not to be frustration reducing and hence not reinforced, were lowered during the course of frustration.

The question may very legitimately be raised: are these results due not so much to the reduction of the secondary frustration-induced instigation which is provided by aggressive behavior as to the more primary reinforcement of aggressive behavior provided by "winning over the adult?"

In Groups A and C, where *O*s took running notes, data exist to provide a partial answer to this question. Such notes were not taken for Group B. Taking cases where all *O*s agreed that such a successful child solution had occurred one or more times, as compared with not at all (12 and 15 cases respectively in the two groups combined), mean aggressive score change was .63 for children having no successful solution as compared with .05 score point for those having one or more successful solutions. The t for this difference is .399, obviously nonsignificant. If we make the assumption that any annotation of successful problem solution, regardless of whether other judges agreed on it or not, was more likely to constitute than not to constitute a successful solution, those children (15 cases) with no successful solutions showed a mean aggressive score change of .63 score point; those with one or more successful solutions showed a mean aggressive score change of 2.89 points (27 cases). The t for this difference is 1.361, again nonsignificant, although in the direction of a "reinforcement hypothesis."

Suggestion that Ag and Su behavior is incompatible was also provided. When scores for the one increase, scores for the other are likely to decrease (r between changes in these two behavior dimensions for the total of 63 cases was −.33, significant at slightly less than the .01 level). Even more strikingly, the correlation for all *S*s between total Ag and total Su was −.651, significant at substantially below the .10 level.

Such results tend to support the notion that even very young individuals have certain characteristic modes of dealing with problems with which they are confronted (personal styles). This notion is also supported by the relationships between teachers' independent ratings of children for LA/n and for PD/n. It might be predicted that such needs (or the behaviors which teachers call to mind in defining such needs) would be incompatible. Such was true for the two rated groups in this study. Older children (Group A) showed an r between LA/n and PD/n of −.75; younger children (Group B) of −.57, both r's being significant at less than the .01 level. The

pooled r was $-.67$, also highly significant. Parenthetically, it is suggested that perhaps characteristic modes of response are strengthened with age, since this relationship as well as all other relationships for Group A showed more stability and strength than was true for Group B.

While the results of the study show ratings of children's social need strengths to be reasonably reliable (greater reliability for PD/n than for LA/n), there is nothing here to indicate that teachers are not defining these needs primarily in terms of behaviors which might be classed respectively as Ag and Su, and that the predictive efficiency of their ratings was due to their using the same base as was used in scoring behavior in the frustration situation.

However, this does not detract from the usefulness of the relationships found. These support the idea that rather broad ratings can be used to predict differential behavior and behavior changes in a frustration situation (or conversely, that the short and convenient frustration task used can be used to predict more generally). In this sense, it appears useful from a prediction standpoint to continue to consider such social-need concepts as love-affection and power-dominance.

Finally, and of tangential interest, it was found that children who had attended permissive preschools for eight months or longer showed significantly higher Ag and lower Su scores during frustration than children who had not had this preschool experience. Preschool attendance seemed to make little difference as far as nature of changes of behavior during frustration was concerned. The authors consider that this supports a social learning notion of frustration-evoked behavior entirely consistent with the original Yale theories.

SUMMARY AND CONCLUSIONS

Preschool children, placed in an eight-trial frustration task, showed significant increases in aggressive-dominant (Ag) behavior scores from the first four to the last four trials, and showed decreases in submissive-complaisant (Su) behavior scores.

Independent raters were able to rate these children with moderate to high reliability for strength of love-affection (LA/n), and power-dominance (PD/n) needs. Children high in PD/n showed higher total Ag scores during frustration, and increased their Ag scores from the first to the second half of the frustration series more than children ranked low. They also showed lower total Su scores and tended to drop such behavior more rapidly than children ranked low in PD/n (but at a level just missing the .05 level).

Children high in LA/n showed smaller total Ag scores and increased these scores during frustration less markedly than children ranked low in LA/n. They also demonstrated higher total Su scores, but did not differ in Su score change from children ranked low on LA/n (in terms of a significant correlation between LA/n and change in Su score).

Ratings for LA/n and PD/n based on CAT protocol ratings were unreliable and entered into no consistent or meaningful relationships with other variables used in the study.

LA/n and PD/n as measured were negatively correlated with each other at a significant level, as were both total Ag and total Su scores and changes of these scores during frustration. Boys did not differ from girls in their patterns of score change, but in the frustration situation showed higher total scores for Ag and lower total Su scores than did girls. Teachers did not rank boys significantly different from girls in terms of either LA/n or PD/n.

Finally, children who had been long-time attenders at a permissive experimental preschool showed higher Ag scores and lower Su scores than children who were newcomers to such a school.

REFERENCES

1. BELLAK, L., & BELLAK, SONYA S. *Children's Apperception Test.* New York: C.P.S. Company, 1949.
2. DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.
3. EDWARDS, A. L. *Experimental design in psychological research.* (3rd Ed.) New York: Rinehart, 1950.
4. LERNER, E., & MURPHY, LOIS B. (Eds.) Methods for the study of personality in young children. *Monogr. Soc. Res. Child Develpm.*, 1941, **6**, No. 4.
5. LINDQUIST, E. F. *Statistical analysis in educational research.* New York: Houghton Mifflin, 1940.
6. MILLER, N. E. (with the collaboration of SEARS, R. R., MOWRER, O. H., DOOB, L. W., & DOLLARD, J.) I. The frustration-aggression hypothesis. *Psychol. Rev.*, 1941, **48**, 337–342.
7. SNEDECOR, G. W. *Statistical methods.* (3rd Ed.) Ames, Ia.: Iowa State Coll. Press, 1946.

Received February 15, 1954.

THE INFLUENCE OF STRESS ON THE PERCEPTION OF INCONGRUITY[1, 2]

CHARLES D. SMOCK

Iowa Child Welfare Research Station, State University of Iowa

INTEREST in the concept of "intolerance of ambiguity" has led to an increasing number of investigations concerned with its psychological correlates (1, 6, 7, 12, 13) but little attention has been directed toward determining the antecedents of the phenomenon. The results of a previously reported experiment by the writer (14) demonstrated that subjects (*Ss*) under psychologically stressful conditions and confronted with ambiguous task situations tended to (*a*) structure stimulus elements into a complete figure relatively earlier in a task series (premature closure), and (*b*) achieve correct recognition of the stimulus relatively later in the series. The first finding was interpreted as lending support to the hypothesis that intolerance of ambiguity, as manifested in the need for cognitive certainty, is a reflection of increased striving for structure and "constancy" resulting from psychological stress.

The relatively retarded recognition of the stimulus was, of course, not directly interpretable from this conceptual frame of reference. One possible hypothesis is that "prerecognition hypotheses" (expectancies) per se interfere with perceptual accuracy (2, 8, 16). However, it was found that the "security" group tended to make as many prerecognition responses as the stress group on the later experimental tasks but at the same time their perceptual accuracy increased (14). It was therefore concluded that prerecognition hypotheses interfere with perceptual accuracy *only* when anxiety is concomitant with those responses. This interpretation fits the data obtained, and at the same time indirectly supports the second predicted consequence of psychological stress in an *ambiguous situation*; i.e., a tendency to adhere to basic or prevailing expectancies (14).[3]

If this assumption is true then individuals under conditions of stress should fail to perceive those elements in an ambiguous perceptual field that are not essential to the "expected," or, in Hebb's (9) sense, to the "familiar" aspects of the field. More specifically, in terms of the experimental tasks, such conditions should lead to a relatively retarded recognition of *incongruous stimulus configurations*.[4] A second hypothesis is that experience in the experimental situation will result in more rapid recognition of incongruity for individuals in both stress and nonstress groups.

PROCEDURE

A more complete description of the experimental procedure is presented in a previous report (14). Briefly, 80 *Ss* were assigned randomly to the experimental conditions and to one of five randomly selected orders of presentation of the five experimental incongruity tasks. An attitude scale administered after the experimental session indicated that the efforts of the experimenter to induce psychological "stress" were reasonably successful.

Incongruity, by definition, indicates a stimulus configuration composed of elements that conflict with the "expectancies" of the individual in the sense he seldom, if ever, is presented with such a stimulus configuration in a "real-life" situation, e.g., half-man, half-woman figure. Five such stimulus tasks were constructed.[5] Each slide was presented to the *Ss* by a systematically controlled series of 22 steps from a position completely out of focus until recognition of the incongruity occurred. The tasks were individually administered immediately following Experiment I, and the *Ss* were

[1] Adapted from the writer's dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology in the graduate school of Syracuse University, 1953. A more detailed report of the experimental procedures and results are presented in the original manuscript which is obtainable from Syracuse University through interlibrary loan service.

[2] The author extends special appreciation to Dr. George G. Thompson for his guidance in the original study, and critical reading of the present paper.

[3] This seems to be the essence of the concept of "rigidity." Brown (3) in his investigation of the relationship between rigidity and authoritarianism modified Rokeach's definition to read: "inability to restructure a field when a familiar structuring is perceived as warding off personal failure." This definition is more consistent with the formulation and data of this study.

[4] The use of incongruity stimuli was suggested by Helson's (10) and Bruner and Postman's (4) studies on the perception of incongruity.

[5] The five stimuli were: half male, half female; old woman and young man; a white man and Negro woman in suggestive positions; and two incongruous playing cards.

instructed to write down, for each presentation, "everything you see or think you see." The criterion of recognition was an unbroken series of three correct responses. A few *Ss* failed to recognize the incongruity on trial 22 and were given one-second exposures at best focus until the correct response did occur.

RESULTS

The data relevant to the hypothesis that stress will result in relatively retarded recognition of incongruity are presented in Table 1.

The difference between the groups on the scores of recognition of incongruity must be interpreted in light of the significant chi square indicating a lack of homogeneity of variance among the groups. However, according to recent empirical evidence reported by Lindquist (11), doubling the confidence yields a fairly reliable and cautious estimate of the probability of obtaining these mean differences irrespective of the heterogeneity of variance. The group effect then is of borderline statistical significance (between $<.10$ $>.04$). This finding lends support to the hypothesis that incongruous elements in the perceptual field are less easily recognized under stressful conditions.

The second hypothesis stated that familiarity with the tasks should result in more rapid recognition of incongruity. The statistical significant "trial" effects indicates systematic variation among the experimental series. In order to test the hypotheses more directly a difference score was computed between the performance on series one versus that of series five. The mean difference for the total group was 1.98 ($t = 4.37$; $p = <.01$). Analysis of the groups separately yields t's of 4.25 ($p = <.01$) for the stress group and 2.05 ($p = <.05 >.02$) for the security group. Figure 1 graphically illustrates this effect of the number of experimental presentations on perceptual accuracy.

The difference between the groups on trial one is statistically significant (diff. 1.40; t 2.2094; $p < .05 > .02$); the difference between the groups is not significant on trials four and five. These results support the hypothesis that intolerance of ambiguity, as manifested by efficiency in recognizing perceptual incongruity, would tend to decrease as a function of the serial presentation of the incongruity tasks. This finding is consistent with the results of Bruner and Postman (4) on the effects of "expectation" for incongruity on the speed of recognition of incongruity.

TABLE 1

ANALYSIS OF VARIANCE OF SCORES FOR TRIAL OF CORRECT RECOGNITION OF INCONGRUITY COMPARING STRESS AND SECURITY GROUPS ON THE FIVE EXPERIMENTAL INCONGRUITY SERIES[1]

Source of Variation	Sum of Squares	*df*	Mean Square	*F*	*p*
Between groups	67.25	1	67.25	5.39	$<.03 >.02$
Residual between individuals	971.51	78	12.45		
Total between individuals	1038.76				
Trials	222.06	4	55.51	4.55	$<.01$
Groups × Trials	23.39	4	5.84		
Pooled error	3800.95	312	12.18		
Total within individuals	4046.40	320			
Total for experiment	5085.17	399			

[1] The Bartlett test for homogeneity of variance yielded a chi square for groups of 14.6134 (corrected) which is significant at less than the .01 level of confidence for one degree of freedom.

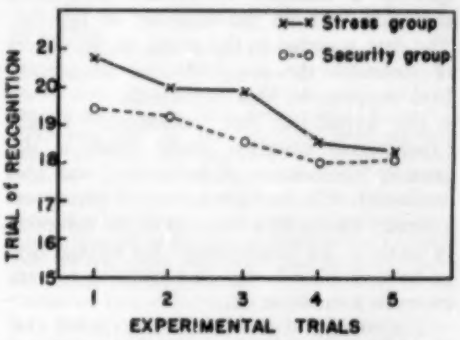

FIG. 1. COMPARISON OF THE MEAN TRIAL OF RECOGNITION OF INCONGRUITY FOR THE STRESS AND SECURITY GROUPS ON THE FIVE EXPERIMENTAL INCONGRUITY SERIES

Correlational data. In the first study (14) it was suggested that the discrepancy score might be considered the most sensitive indicator of intolerance of ambiguity on the cognitive closure task inasmuch as it summated the effects of stress on both the speed of closure and the predicted interference effect of prerecognition response delaying correction recognition. The degree of correlation between this discrepancy score and recognition of incongruity score would give an indication of the extent to which both tasks tap psychological processes related to intolerance of ambiguity. The obtained correlation (+.43) is significant beyond the .001 level of confidence.

The trial of correct recognition (cognitive-closure task) and recognition of incongruity were not significantly related ($r = +.17$); neither were the speed of closure and the recognition of incongruity ($r = -.15$). Both correlations are in the predicted direction, however. An experiment specifically designed to study the degree of relationship of the two variables might be expected to produce more definitive results.

SUMMARY AND CONCLUSIONS

The major hypothesis of this study was that individuals under conditions of psychological stress will become more intolerant of ambiguity. Specifically, it was hypothesized that psychological stress will lead to the adherence of prerecognition "hypothesis" and consequently a retarded recognition of stimuli that deviate from the expected or familiar. The data reported in this study on the effects of stress on the recognition of incongruity lend support to this hypothesis.

The hypothesis that "experience" in the experimental situation would result in decreased "intolerance of ambiguity" was also confirmed. Whether the increased perceptual accuracy was more a function of the reduction of anxiety, an "expectancy" for incongruity, or both of these factors is a question not answerable from these data.

The results of the initial experiment (14) and the one reported here support the hypothesis that psychological stress results in (*a*) premature closure and (*b*) a tendency to adhere to expectancies (prerecognition hypothesis) in ambiguous task situations. From these two sets of data it is inferred that stress or anxiety results in cognitive and perceptual processes that tend to preserve a "familiar" perceptual and behavioral field for the individual.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BLAKE, R. R., & VANDERPLAS, J. M. The effect of prerecognition hypotheses on veridical recognition thresholds in auditory perception. *J. Pers.*, 1950, **19**, 95–115.
3. BROWN, R. W. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.
4. BRUNER, J. S., & POSTMAN, L. On the perception of incongruity: a paradigm. *J. Pers.*, 1949, **18**, 206–223.
5. EDWARDS, A. *Experimental design in psychological research*. New York: Rinehart, 1950.
6. FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual variable. *J. Pers.*, 1949, **18**, 108–143.
7. FRENKEL-BRUNSWIK, ELSE. Personality theory and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald Press, 1951. Pp. 356–420.
8. GALLOWAY, D. W. An experimental investigation of structural lag in perception. *Amer. Psychologist*, 1946, **1**, 450. (Abstract)
9. HEBB, D. O. *Organisation of behavior*. New York: Wiley, 1949.
10. HELSON, H. The relation between instructions and past experience in a simple observational task. *J. educ. Psychol.*, 1934, **25**, 29–38.
11. LINDQUIST, E. F. *Design and analysis of experiments*. Cambridge: Riverside Press, 1953.
12. O'CONNER, PATRICIA. Ethnocentricism, "intolerance of ambiguity," and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526–530.
13. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentricism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
14. SMOCK, C. D. The influence of stress on the "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, **50**, 177–182.
15. WYATT, D. F., & CAMPBELL, D. T. On the liability of stereotype or hypothesis. *J. abnorm. soc. Psychol.*, 1951, **46**, 496–500.

Received February 23, 1954.

A PROPOSED METHOD FOR INCREASING ACCURACY OF JUDGES' CLASSIFICATION OF VERBAL MATERIAL[1]

S. ROSENBERG AND B. I. LEVY

Air Force Personnel and Training Research Center, Randolph Air Force Base, Texas

MANY research problems in psychology require the use of systems of content analysis as a prerequisite for quantitative treatment of data. These systems (2, 6, 8, 9, 11, 12) require the use of judges who classify verbal units within one of several categories. The "relative" or "absolute" reliability (10) among judges is frequently the focus of considerable attention. A number of factors may be proposed which contribute to the difficulties of obtaining adequate reliability. One factor which may be involved is the arbitrary and abstract nature of the classification system. While a research worker who originates a content analysis may score reliably, himself, the use of the system by other judges is not easily accomplished. Typically, an elaborate set of instructions and training for judges must be developed so that judges will respond to the content in a "standard" manner. Studies of how instructions and training are most effectively developed from abstract definitions are generally lacking. The present paper is concerned with a systematic method for developing a set of definitions that will increase judge accuracy, i.e., conformance to a standard of performance.

METHOD

Subjects (Judges)

Two groups of forty-eight male subjects (*Ss*) each were used as judges in the study.[2] Thirty-four *Ss* were airmen assigned to the Air Force Personnel and Training Research Center as research assistants. Sixty-two other *Ss*, who volunteered for the experiment, were college students at Trinity University, San Antonio, Texas, in freshman, sophomore, and junior ROTC classes. The distribution of education among the airmen is approximately the same as that of the ROTC students. Random assignment was used for the two groups except for the restriction that the airmen and ROTC students be equally represented in each group. The *Ss* had no knowledge of the general purpose of the experiment except for the information furnished by the instructions.

Material Judged

All *Ss* were given a total of 300 clauses to classify according to a content system proposed by Dollard and Mowrer (2)[3] and subsequently employed by others (1, 3, 5, 7). Two separate booklets, containing 150 clauses in each booklet, were prepared for this purpose. The instructions and definitions of the D, R, and N categories appeared on the first page of each booklet. Separate IBM sheets were provided for *Ss* to record their judgments.

The clauses were selected from written Thematic Apperception Test (TAT) protocols of bomber crew members who had taken the TAT in connection with another study. In that study four questions were asked about each card: (*a*) Who are they? What are they doing? (*b*) What has happened? (*c*) What are they thinking? (*d*) What will happen? Protocols from four cards (1, 6BM, 8BM, and 10) of 19 examinees were sampled so that 16 statements were obtained from each person. As much as possible, the statements were taken from the first clause responses to each question of each card except for the questions, "Who are they?" and "What are they doing?" The second clause was used here, instead, since the first clause in response to these questions was generally too descriptive to provide adequate sampling from the three categories.

This procedure of clause selection yielded 304 statements in all. In order to adjust the number of clauses to the IBM answer sheets four clauses were discarded. Clauses were arranged in the booklets in random order.

Procedure[4]

Phase 1 (Preliminary Judgments)

The two booklets of clauses were given so that half the *Ss* worked with one booklet, A, first, while the other

[1] This research was carried out under the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas, in support of Project No. 7713. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government.

[2] The experiment required that data be collected on a "test-retest" basis. There were 114 *Ss* who were present for the first session. Eighteen *Ss* were absent for the second session, leaving 96 *Ss* for statistical analyses.

[3] The analysis of content proposed by Dollard and Mowrer requires that a protocol first be broken down into clauses and then classified. Three categories of classification are defined. Units expressing "drive, want, tension, pain, suffering" are classified as Discomfort (D); those units which express "drive reduction, reward, pleasure, relaxation" are scored as Relief (R); and finally, those units expressing "neither discomfort nor relief," or expressing "both in approximately equal amounts" are scored Neutral (N).

[4] We wish to acknowledge the assistance of S. Sgt William Keown, A/1C Milo Peterson, A/2C Richard Mourglia, and A/2C Harlan Hogsven in the data collection and statistical analysis.

half worked with the other booklet, B, first. When an *S* finished the first booklet, he was given the second booklet. This counterbalancing was done in an attempt to distribute practice and fatigue effects over all items.

The *Ss* were told that they were judges in an experiment—*not examinees*. They were instructed to place each item (clause) into one of the three categories according to the written definitions in the booklet. These instructions as well as the definitions were read to *Ss* in both groups by *E* exactly as they were written in the booklet.

The definitions given to the *Ss* are similar to those originally provided by Dollard and Mowrer (2). Minor changes were made in the use of technical psychological language, e.g., substitution of "want" for "drive," and "release . . ." and "removal . . ." for "drive reduction." No other instructions or training were given.

Phase 2 (Definition Revision)

Since both groups were treated in the same way in Phase 1, the judgments of all 114 *Ss* were massed for item counts.[5] For each clause the number of D, R, and N responses was counted. The category in which the modal frequency occurred determined the assignment of the items as a "D item," "R item," or "N item." Of the 300 items 142 were found to have modal frequencies in D, 90 in R, and 68 in N. Within each category items were then arranged in rank order according to the size of the mode.

Both *Es* read through the ranked items within each category and then attempted to formulate more explicit definitions or principles for the categories. In reading through the clauses it was evident that items within a category could be grouped according to fairly discrete areas of content. Ranking the items according to the mode revealed, in many cases, that items with the larger modes were the most obvious examples of a content area, while items with smaller modes provided more subtle examples of the same area. Some items, particularly those with small modes, appeared sufficiently heterogeneous so that no grouping of them could be made.

Both *Es* read through each item separately and then in unison until all items were agreed upon as belonging to a particular subcategory within one of the three major categories. Items which could not be assigned to a subcategory were omitted from further consideration. After grouping the items, the *Es* composed descriptions for each grouping that were estimated to be comprehensible to the *Ss*.

The aim of the definition-making procedure was to provide explicit definitions without using the words contained in any of the clauses in formulating a definition. The following "new definitions" for each category were written:

Discomfort category. Statements which describe or imply that a person(s) (also animals and objects) experiences one of the following: (*a*) separation of close personal relationships: conflict between people; (*b*) statements of sorrow, mental suffering, fear, fatigue, boredom; (*c*) failure or expected failure to achieve a goal regardless of its importance; (*d*) being helpless in problem situations: unable to make up one's mind; (*e*) striving of any kind by a person(s). Statements which describe or imply that a person(s) (also animals and objects) receives or is exposed to one of the following: (*a*) physical (bodily) damage such as death, injury, illness, operations; (*b*) threat of physical damage; (*c*) report of misfortune, minor or major; (*d*) disgrace; (*e*) being kept from doing something: being compelled to do something.

Relief category. Statements which describe or imply that a person(s) (also animals and objects) receives or experiences one of the following: (*a*) physical or spiritual love relationships; other warm, cordial feelings, thoughts, or actions among people; (*b*) pleasure, happiness or comfort coming from desired objects or happenings; (*c*) anticipation or achievement of any planned or desired goal, regardless of its importance. Statements which describe or imply escape, avoidance, removal, or termination of something undesirable: (*a*) people coming together again after separation; people making up with each other; (*b*) recovery from injury; stopping of punishment, threat, or fear of any kind.

Neutral category. Statements which clearly do not imply discomfort or relief as defined above. These neutral statements simply tell of: (*a*) existence of any person(s), objects, or events not previously defined as discomfort or relief; (*b*) the existence of mental or physical activity not previously defined as discomfort or relief. In addition, the following general rules were thought to be helpful as supplementary instructions for an *S*:

1. Read both past and future references as though they are happening in the present. For example: I was . . . , I will . . . , I am . . . , are all to be read as "I am . . . "

2. Do not allow kind or type of person to prejudice your classification of the statement. In other words, consider "I," "you," "they," "males," "females," "doctors," "lawyers," "children," etc. as nothing more than people. Remember, the entire statement is to be used as a basis for a judgment.

These two rules were added when it became apparent that a secondary loading in one of the nonmodal classifications was associated with the tense of the verb in the clause and/or the cultural connotation of the subject of the clause. The rules are an attempt to minimize these biases. Rule 1 appears in Dollard and Mowrer's directions (2).

Phase 3 (Retest)

The same 300 clauses were readministered to all *Ss* approximately nine weeks after Phase 1. The *Ss* in Group I were provided with the same definitions they

[5] An 11-page table giving the sample of 300 clauses and the frequencies within each category obtained from the 114 *Ss* in the first phase of this study has been deposited with the American Documentation Institute. Order Document No. 4337 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.75 for 35 mm. microfilm or $2.50 for 6 × 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

had used the first time. The *Ss* in Group II were given the new definitions, and also were given the two general rules as supplementary instructions. All other instructions remained the same.

Counterbalancing of booklet order was also done in Phase 3. In addition, the order of the booklets (A and B) was reversed for each *S* from the order he had used in Phase 1. Thus, the *Ss* who had had A-B order in Phase 1 were now given B-A order, and vice versa.

RESULTS

In evaluating the effectiveness of the new definitions, it is necessary that the essential characteristics of the original system be retained. Thus, a change in definitions ought not to result in a shift of the item to another category. It would be a simple matter, without such a limitation, to get high agreement among *Ss* by having them categorize clauses according to whether or not they contained specific words. By assigning an item to one of the three categories according to that item's modal frequency in Phase 1, and, furthermore, by evaluating any changes from this standard, it is assured that the characteristics of the original system are retained.

Once the standard category for each item had been specified, three scores were obtained for each *S*, i.e., the number of correct D, R, and N responses. Since there were 142 D items, 68 R items, and 90 N items, these figures represent the maximum correct scores.

Table 1 contains a summary of the data from each group of *Ss*. The means for Phase 1 are computed for each of the three sets of items as well as for the total number of correct responses for all 300 items. The means of the difference scores, obtained by subtracting the second administration from the first, are also shown in Table 1.

These data indicate an over-all gain in the number of correct responses in the condition containing the new definitions. Examination of the three sets of items separately, however, reveals gains for the D and R items but a loss for the N items. To assess the significance of these observed trends, the difference scores were subjected to an analysis of variance. The statistical design involved two major criteria of classification, i.e., the 96 individuals and the three categories. Individuals were further subdivided into the two groups. From

TABLE 1

MEAN FREQUENCY OF CORRECT RESPONSES FOR THE D, R, AND N ITEMS SCORED SEPARATELY AND ALSO COMBINED

($N = 96$)

Category	Group I		Group II	
	Phase 1	Phase 3* − 1	Phase 1	Phase 3* − 1
D items (142)	111.04	−6.29	111.19	5.85
R items (68)	49.02	−2.00	50.58	3.11
N items (90)	53.96	2.83	56.85	−2.25
All items (300)	214.02	−5.46	218.62	6.71

* Phase three means are omitted to conserve space. They may be calculated from data given in the table.

this design[6] the following sources of variance were obtained:

1. Categories: D, R, N;
2. Individuals
   - *a*. Between Groups: I, II
   - *b*. Pooled Within Groups;
3. 1 × 2
   - *a*. 1 × 2*a*
   - *b*. 1 × 2*b*.

A significant statistical difference, beyond the .05 level, in the expected direction, between Groups I and II (source 2*a*) was found using source 2*b* as denominator ($F = 5.31$). The results for this main effect support the contention that the new definitions have contributed to increased accuracy.

The interaction of categories with groups (source 3*a*) was also tested using source 3*b* as divisor, yielding an F ratio beyond the .01 level ($F = 5.38$). This statistical result is not surprising in view of the reversal found for the N category means as compared with the D and R category means when the new definitions are used. In view of the significant effect between groups, however, it appears that the gains in the D and R categories are significantly greater than the loss in the N category.

To assess the trends in each of the categories, separately, three CR's were computed from the same data. With the difference scores, the two groups were compared on the D category items. The CR between groups was 2.77, significant beyond the .01 level. Simi-

[6] The general model for this analysis may be found in an article by Kogan (4).

larly, the comparison of the two groups on the R category items yielded a CR of 2.09, significant beyond the .05 level. Finally, the comparison on the N items yielded a nonsignificant CR value of 1.52.

The discrepancy from what was expected in the N category, although not significant, led to some further examination of the data. It is possible, for example, that the new definitions have resulted in "sensitization" to the D and R categories; i.e., *Ss* make more D and R responses and fewer N responses. If so, the gain in number of correct responses in D and R items may simply be the result of more D and R classifications, per se, rather than more accurate discriminations. Thus, for the D items, the number of R classifications may have *increased also*. This hypothesis was tested in Group II by examining the change in frequency count in the nonmodal R category with the 142 D items and the change in the nonmodal D category with the 68 R items. Actually, there was a mean decrease for Group II in both instances (nonmodal R and D categories), although not significant when examined with CR's using the difference scores. Therefore, the increase in judge accuracy cannot be attributed to indiscriminate sensitization to the D and R categories.

DISCUSSION

Although the results of the present study are generally significant, several points require further consideration. One of these is the fact that the mean frequencies for the D and R categories do not demonstrate extraordinarily large gains. Thus, despite the gain that does occur, 17.6 per cent of the responses to the D items continue to be misplaced by Group II. In the R category, Group II *Ss* misplaced 21.0 per cent of their responses; and for the N category, 39.3 per cent misplacement occurs. These proportions of "error" appear to be much larger than are obtained with mechanical instrumentation involving frequency counts. Additional instruction and/or training of judges, which would attempt to reduce "error," must, of course, also conform to the specification of reproducibility among independent experimenters.

A second point is concerned with the choice of judges. It seemed acceptable to have research technicians and/or any comparable population provide the preliminary classification of items on the basis of which revised definitions would be proposed. This choice was based on the assumptions that (*a*) initial discriminations made by this population would reflect some of the limits that the use of populations without psychological training impose, and (*b*) nonprofessional persons have had cultural and idiosyncratic histories which provide them with *some* discriminatory behavior concerning discomfort and relief statements.

The brevity of the definitions for making consistently adequate judgments was counteracted by the use of a relatively large number of judges so that modal or majority judgments were assumed to be a sufficiently stable and valid representation for an item.

A third point is the possibility that the definitions derived from the categorization of the 300 items are not generalizable to a sample of new items from verbal productions in situations other than the TAT. The procedure used by the experimenters in constructing new definitions included attempts to avoid defining principles containing the specific words in the actual clauses to be judged. Whether this procedure minimized the development of principles applicable only to the 300 items is a matter for further study.

Another point concerns the reproducibility of the revised definitions. The question here is whether other experimenters, given the preliminary classification data, would have arrived at similar principles, using the procedure described for Phase 2. This problem may be studied empirically by providing situations where as many experimenters ("definition-makers") as seemed desirable could participate independently and/or jointly in the definition construction. "Similarity" of definitions by these independent experimenters may be assessed by comparing judges' conformance to the standard.

A final issue is related to the precise manner in which the new definitions have made their contribution to the obtained results. For example, can the shift in accuracy be attributed to the new definitions or to the new definitions plus some concomitant factor? It is possible that the effectiveness of the new definitions was contingent upon previous practice with the item, even if this practice

involved the use of the old definitions. Data relevant to this point were obtained during the retest of the college students. A sample of 62 new *S*s was made available from the same classes as the *S*s in the first experiment. The new sample consisted of students who had not volunteered originally. These new *S*s were split into two groups and given the old and new definitions, respectively. Since the method of sampling these *S*s was not comparable to that used originally, the test of relative efficacy of the new versus old definitions is fortuitous. In general, the results with this sample were similar to those obtained in the experiment proper, but statistical tests failed to yield significance. A more intentional design than the preceding would be needed to isolate the relationships between kind of practice and accuracy of judgment.

Although the procedure described in this paper was developed for the D-R-N system, it is proposed that the method may be employed with other systems where complex behavior categories are desired. Furthermore, these behavior categories need not be restricted to written material but could possibly be extended also to auditory and/or visual samples with the aid of recording devices.

In attempting to summarize the general methodology in this study, it appeared profitable to employ a concept-formation analogue.

As is true in most concept-formation experiments, *S*s in this study are asked to place stimuli possessing common characteristics in categories. In contrast to other concept studies, however, in this study positive and negative instances of membership in a category were not explicitly indicated; moreover, *S*s were not told whether they were right or wrong. Instead, they were provided with the defining features of each category. The concept-formation paradigm also suggests that an initial set of definitions is not necessary for obtaining defining characteristics of a class of stimuli. A worker may simply wish to submit behavior units and indicate to the subjects which are positive and negative instances of a stimulus class. The subjects may then be asked to verbalize the definitions of positive and negative instances.

Finally, an additional procedure, suggested by the concept-formation analogue, may be appended to the method employed here. Judges can be given sample items to classify using the new definitions and would be told after each response whether they were "right" or "wrong." Such a procedure may reduce error considerably since it was observed that most judges did not refer back to the definitions on the first page of the booklet in categorizing items on subsequent pages. The gains which were found seem to be largely attributable to the initial reading of definitions. Use of the additional training procedure of indicating correct and incorrect responses would "force" judges to become better acquainted with the principles and would probably lead to an increase in the accuracy of judges' classifications.

SUMMARY AND CONCLUSIONS

Many research problems in psychology require the classification of controlled or spontaneous verbal responses. Typically, because such classification systems are not explicitly defined, they necessitate the use of sophisticated or extensively trained judges to insure adequate reliability. This paper contains: (*a*) a method by which category definitions are derived from preliminary classification by nonprofessional judges, and (*b*) tests of the effectiveness of these definitions for increasing judge accuracy.

A sample of 96 judges, consisting of airmen and college ROTC students, was asked to classify 300 clauses taken from TAT protocols. The classification system was based upon the discomfort, relief, and neutral categories described by Dollard and Mowrer. On the basis of these results, items were placed in the categories according to the modal frequency of the judgments. From these groupings the authors developed more explicit content principles for each category. New definitions were based on these principles. The 300 clauses were then readministered to the sample with half the judges using the original definitions and the other half using the new definitions.

Several analyses were performed to determine if the number of correct judgments was increased through the use of the new definitions. The more prominent results indicated: (*a*) An over-all increase in judges' accuracy of classification in the new-definition group which was significantly different from any

changes in the old-definition group, and (*b*) a differential increase in judge accuracy among the categories, with discomfort and relief increasing while the neutral category shows a decrease.

It is concluded that the proposed method can be utilized for increasing interscorer accuracy. The possibility of employing the method for other classification systems is considered.

REFERENCES

1. ASSUM, A. L., & LEVY, S. J. Analysis of a nondirective case with followup interview. *J. abnorm. soc. Psychol.*, 1948, **43**, 78–89.
2. DOLLARD, J., & MOWRER, O. H. A method of measuring tension in written documents. *J. abnorm. soc. Psychol.*, 1947, **42**, 3–32.
3. JOHNSON, JOANNE E. The relationship between discomfort-relief quotients and scores on the BI-N (neurotic tendency) scale of the Bernreuter Personality Inventory. Unpublished master's thesis, Pennsylvania State Coll., 1950.
4. KOGAN, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.
5. KOGAN, L. S. The distress-relief quotient (DRQ) in dictated and verbatim social casework interviews. *J. abnorm. soc. Psychol.*, 1951, **46**, 236–239.
6. MCCLELLAND, D. C., CLARK, R. A., ROBY, T. B., & ATKINSON, J. W. The projective expression of needs: IV. The effect of the need for achievement on thematic apperception. *J. exp. Psychol.*, 1949, **39**, 242–255.
7. MEADOW, A., GRLENBLATT, M., LEVINE, J., & SOLOMON, H. C. The discomfort-relief quotient as a measure of tension and adjustment. *J. abnorm. soc. Psychol.*, 1952, **47**, 658–661.
8. MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
9. RAIMY, V. C. Self-reference in counseling interviews. *J. consult. Psychol.*, 1948, **12**, 153–163.
10. THORNDIKE, R. L. *Personnel selection*. New York: Wiley, 1949.
11. TOMKINS, S. S. *The Thematic Apperception Test*. New York: Grune & Stratton, 1947.
12. WHITE, R. K. Black boy: a value analysis. *J. abnorm. soc. Psychol.*, 1947, **42**, 440–461.

Received February 26, 1954.

A FIELD EXPERIMENT ON RUMOR TRANSMISSION AND DISTORTION[1]

STANLEY SCHACHTER AND HARVEY BURDICK

University of Minnesota

RUMOR is usually characterized as an unreliable, sometimes wildly distorted form of communication which spreads rapidly and mysteriously to almost all available members of a population. The conception of distortion and exaggeration as characteristic of such forms of communication arises largely from generalization of findings in studies of perception and memory using the technique of serial reproduction (1, 3). The impression of widespread and rapid diffusion is, with the exception of numbers of dramatic anecdotes, relatively undocumented, for there have been few studies of the spread of rumor.

The results of the few systematic studies which have been published are peculiarly at variance with the conception of rumor outlined above. In Table 1 the results of four such studies are outlined. In the first three of these studies (2, 6, 8), rumors were planted and then by means of systematic interviewing or participant observation of the relevant population, it was possible to make reasonably accurate estimates of the extent of rumor spread. For the 13 planted rumors reported in these three papers, an average of only 5 per cent of the relevant population per rumor actually heard the rumor. An average of about 3 per cent per rumor is reported as having told the rumor. In addition, the authors are familiar with two other studies using planted rumors; these have remained unpublished because of complete failure of the rumors to spread. The present paper is prompted, in part, by sheer exasperation at the growing number of such abortive studies.

Whether such results should be attributed to experimental failure or accepted as evidence that our notions about rumor are much exaggerated is equivocal. Comparison of these studies with reports of spontaneously arising rumors is difficult, for so few of the studies concerned with spontaneous rumors present any accurate estimate of the extent of rumor spread. One exception is the study of Festinger, Cartwright, *et al.* (item 4 in Table 1) of a spontaneous rumor concerning communist activity in a housing project. Fifty-nine per cent of the interviewed population reported having heard the rumor, 21 per cent reported having told the rumor—considerably greater percentages than those reported in any of the studies employing planted rumors. However, since this single study was conducted specifically because the authors' attention was drawn to an already widespread rumor, it is impossible to make any guesses as to the extent to which such results may be typical of spontaneously arising rumors.

The limited data available on rumor distortion are similarly at odds with current expectations. In the two studies (2, 6) listed in Table 1 in which the senior author participated, and is, therefore, familiar with data not reported in the published papers, there was almost no evidence of any distortion or modification of the planted rumors. In the Festinger, Cartwright, *et al.* study (5), rumor spread was tested by a recognition question which made it impossible, of course, to evaluate distortion.

Such results as these are so puzzlingly at odds with current notions of the nature of rumor that one is forced to examine critically the studies here reviewed as well as to reexamine our current conceptions of rumor. We will consider first the problem of the spread of rumor and reserve for consideration in a later section of this paper the problem of rumor distortion.

In an attempt to formalize the conditions of rumor spread, Allport and Postman propose as "the basic law of rumor" that "the amount of rumor in circulation will vary with the importance of the subject to the individuals concerned times the ambiguity of the evidence pertaining to the topic at issue" (1, p. 34). A similar proposition is the "principle of cog-

[1] This study was conducted under sponsorship of the Laboratory for Research in Social Relations of the University of Minnesota. The authors wish to thank Marion MacKenzie, Beatrice Tselos, and Marjorie Schachter for their assistance in planning and executing the study.

TABLE 1

RUMOR SPREAD IN FOUR STUDIES OF RUMOR

Study	Number of Rumors	Population × No. of Rumors	Percentage of Population Per Rumor Who Heard Rumor	Percentage of Population Per Rumor Who Told Rumor
Back *et al.* (2)	9	55 × 9	5.9*	4.2*
Festinger, Schachter, and Back (6)	2	100 × 2	5.1	2.5
Schall, Levy, and Tresselt (8)	2	10 × 2†	0	0
Festinger, Cartwright, *et al.* (5)	1	100 × 1	59.0	21.2

* From information reported in the publication, it is impossible to compute the exact figures. The figures reported are maximum possible estimates.

† Rumors were planted in two different groups but N is reported for only one of these groups. It is assumed that the groups were roughly the same size.

nitive unclarity" by Festinger, Cartwright, *et al.* "Rumors will tend to arise in situations where cognitive regions especially relevant to immediate behavior are largely unstructured" (5, p. 484). Both formulations agree in identifying cognitive unclarity or ambiguity and importance or relevance as key determinants of the origin and spread of rumor. The latter formulation is somewhat more specific in its treatment of the "importance" variable, relating it to areas with relevance for immediate behavior. We shall use the term in this sense.

It is probably wise, too, to make explicit what is undoubtedly implicit in both formulations—that the state of cognitive unclarity about an important issue be common to all or a major part of the population under consideration. Our apparent insistence on this factor arises from consideration of the pattern of communication which rumor customarily follows. In most social communication there is a back and forth exchange between two or more people and little further relaying of the contents of such conversations. The form of communication called rumor is characterized by a chain pattern of communication. A communicates an item to B, B communicates the item to C, C to D, and so on. In such a pattern, possession of the item of information seems to create a force to communicate it further. If one presumes that one kind of situation giving rise to such forces to communicate may be characterized by cognitive unclarity about an important issue, it seems clear that these conditions must hold for all or a large portion of the group in order that the rumor spreads. Otherwise, this chain pattern of communication will be quickly interrupted[2] and, unless there are particularly persistent rumormongers involved, there will be little spread of the rumor.

Though none of the studies outlined in Table 1 attempted a direct test of the relationship of unclarity and importance to rumor spread, the technique employed in almost all of these studies assumed an acceptance of these hypotheses concerning the determinants of rumor spread. Usually the experimenters attempted to identify areas which seemed important and about which there may have been some ambiguity and then to plant rumors relevant to these areas. Since none of these studies collected evidence which would allow proper evaluation of the extent to which these conditions were satisfied, it would seem a fruitless sort of post-mortem to examine the details of these studies in terms of the postulated determinants of rumor spread.

Rather than attempt to explain such failures of rumor spread in terms of a conceptualization which is still largely untested, it would seem more worthwhile to proceed to a direct test of the conceptualization. The present study is a field experiment designed to test some of these ideas of the determinants of rumor spread. A situation which plausibly can be described as involving cognitive unclarity about an important issue is experimentally manipulated and rumors relating to the issue are systematically planted.

METHOD

Background. The study was conducted in a girls school. Three months before this study was conducted, two of the teachers in the school decided to conduct a demonstration experiment on rumormongering. The procedure they adopted was similar to that employed in most studies in which planted rumors are used. Two girls were let in on the secret, instructed how to

[2] This seems to be the case with most of the rumors listed in Table 1. In the Festinger, Schachter, and Back study, 71 per cent of all communications were initiated by the 4 per cent of the population with whom the rumors were originally planted. In the Back *et al.* study an estimated minimum of 26 per cent of all communications was initiated by the 4 per cent of the population with whom the rumors were planted.

proceed, and during the course of one day planted the rumor with some 10 to 15 girls.

The rumor concerned the school's Christmas play. There had been some friction between the teacher directing the play and the girls involved and there had been a series of quarrels and petty incidents concerning casting, rehearsals, and the like. The rumor was that the Christmas play had been cancelled. Some of the girls with whom the rumor was planted were involved in the play; others were not.

The results again were nil; there seems to have been virtually no spread of the rumor. The two girls who had planted the rumor and several of the teachers acted as participant observers and were unable to report a single act of communication. As a demonstration, the experiment was an unfortunate failure; but it does serve as a casual additional control for the present study since precisely the same girls were the subject population for the attempt at rumor spreading described in this paper.

Population. The school in which the study was conducted is a small, exclusive, girls preparatory day school. The school is both a primary and secondary school and has twelve grades with one class of about 18 students constituting each grade. The school day begins at 8:25 and ends at 4:00; the girls do not return home during this period. Most of this time is occupied with standard class work, but each grade has a one-hour lunch period and a 15-minute recess at 10:15 during which time the girls meet together in the lunchroom. Most of the girls have a one-hour study period and an hour of gym. During recess, at lunch, and in gym periods the girls are free to communicate with one another. There is a five-minute interval between class periods during which time the girls change classrooms and are again free to communicate and socialize.

The experiment proper was conducted in the upper six grades of the school—the standard junior high school and high school grades. The age range of the girls concerned was between twelve and a half and seventeen and a half.

Design. There were three experimental conditions: (*a*) The "cognitive unclarity-rumor" (CU-R) condition—A situation of cognitive unclarity was manipulated, and a rumor was planted. (*b*) The "cognitive unclarity" (CU) condition—a situation of cognitive unclarity was manipulated and no rumor was planted. (*c*) The "rumor" (R) condition—a rumor was planted, but no situation of cognitive unclarity was created.

Six classes containing a total of 96 students were involved in the study. Two classes were assigned to each experimental condition so that in each condition there was one of the older three classes and one of the younger three classes.

Producing a situation of cognitive unclarity. On the day of the study, between 8:25 and 8:35, the principal of the school went into four different classrooms. In each class, she interrupted the work, stood in front of the class, pointed a finger at one girl, and announced, "Miss K., would you get your hat, coat, and books, please, and come with me. You will be gone for the rest of the day." Then, without a word, she and the girl walked out of the room together. Such an action was completely unprecedented in the experience of the girls. To insure that the event remained a complete mystery, the entire staff of the school had been instructed to reply to any questions about the event that "they knew nothing about it."

The girls taken from each of the four classes had been selected on the basis of their sociometric status and their academic and disciplinary records. All four girls were matched on these three criteria. They were chosen so as to fall between the fiftieth and seventy-fifth percentile in grade average and in the number of sociometric choices they received from their classmates. In addition, none of the girls had a disciplinary record. These girls, then, were fairly average members of their classes, reasonably popular and with fair grades. None of them knew anything about the study before they were taken from class.

Was the manipulation effective in producing the desired constellation of variables? As part of the standard data-collecting procedure, the teachers in the school had all been instructed to keep a record of all questions addressed to them about the manipulation. They were asked to record who spoke to them, when, and the content of the question or remark. There were 62 girls in the four classes from which the principal had removed one student. The teachers reported a total of 198 questions, plus the report of one harried teacher, unable to keep up with the curiosity of the girls, that "everyone in the class asked me what had happened" and of another that "half the class asked questions." Virtually every girl in the classes affected made inquiries of one or more of her teachers and almost all questions were of the sort, "What happened to Miss K.?" "Why did the principal take Miss K. out of class?" "What's going on?" etc. We assume from the nature of these questions, which were largely expressions of curiosity, puzzlement, and attempts to get information, that the manipulation did produce a state of cognitive unclarity common to all or most of the girls involved. Evidence relevant to the importance variable will be presented in a later section.

It should be pointed out that this manipulation has undoubtedly produced more than a simple state of cognitive unclarity. The event was novel and dramatic, and there can be little doubt that it produced surprise and excitement as well as considerable curiosity and unclarity as to what was going on. We have abstracted from this constellation the element of cognitive unclarity, and the nature of the questions asked the teachers would seem to justify this characterization, but it is likely that the effects of the manipulation on rumor transmission are in part attributable to the additional elements involved.

Planting the rumor. The rumor was planted with two girls from each of four different classes. Two of these classes were from the four in which the situation of cognitive unclarity was produced; in the remaining two classes there had been no such manipulation. The eight girls with whom the rumor was planted were also matched in terms of their sociometric, academic, and disciplinary records.

A day or two before the study took place, various teachers made appointments for 8:15 on the morning of the study to see each of the eight girls with whom the rumor was to be planted. Ostensibly, the purpose of these appointments was to discuss academic progress, next year's program, etc. This was routine procedure.

TABLE 2

KNOWLEDGE OF THE RUMOR AND THE COGNITIVE UNCLARITY MANIPULATION

Condition	*N*	Per Cent Knowing of CU Manipulation	Per Cent Knowing Rumor
CU-R1	18	100	100
CU-R2	15	100	100
CU-R1 + CU-R2	33	100	100
CU1	18	94	100
CU2	11	100	100
CU1 + CU2	29	97	100
R1	18	94	94
R2	16	100	100
R1 + R2	34	97	100

Each of these interviews followed an identical pattern. After six or seven minutes of discussing the matter for which the appointment had presumably been arranged, and immediately before terminating the interview, each teacher said, "By the way, some examinations have been taken from the office. Do you happen to know anything about this?" No such thing had taken place and, of course, all of the girls interviewed denied any knowledge of the affair. The interview was so timed that each of the girls returned to her classroom before the principal entered any of the rooms.

The rumor planted in this way was intentionally chosen so as not to be an immediate explanation for the "cognitive unclarity" manipulation. It could, however, be readily linked to the morning's events as a tentative sort of explanation. The rationale for this choice is the following. One of our interests was the origin of new rumors. It seemed preferable to entrust the job of planting the rumor to teachers rather than making confederates of a number of girls and thereby imposing on them the difficult job of maintaining a standard pattern of behavior throughout the day. It seems a reasonable hunch that a clearly explanatory rumor originating from an authority figure, such as a teacher, would tend to inhibit the kind of speculation from which new rumors arise.

Data collected. Three types of data were collected:

1. Sociometric data: Three weeks before the study took place all of the girls involved answered a sociometric questionnaire. The school was planning a school fair for which the girls were to work together at planning exhibits, running shows, etc. The sociometric questionnaire was linked to this affair and read: "As you know, we are hoping to hold the school fair early in the fall. We would like you to work with girls you enjoy being with. If you will list your five best friends we will try to arrange it."

2. Teachers' observations: As previously mentioned, all of the teachers kept a record of comments addressed to them concerning the manipulation.

3. Standardized interview: At two o'clock in the afternoon, toward the end of the school day, a team of 20 interviewers took over the school lunchroom in order to interview all of the girls involved in the study. The interviewing schedule was so arranged that an entire class was interviewed at the same time in order to prevent any communication among the girls about the nature of the interview. The classes were scheduled and brought down to the interview room in a fashion which prevented any communication among classes.

The interview was a standardized, open-end instrument designed to get information about what the girls had heard about the situation created by the cognitive unclarity manipulation, with whom they had talked about this, whether or not they had heard the planted rumor, from whom they had heard it, and to whom they had told it, and how much time they had spent discussing all of these matters during the day.

At the end of the day, after all of the girls had been interviewed, a general assembly was called for all of the classes involved in the study. The study was explained in complete detail at this gathering. The four girls who had been taken from their classes (and who had spent the day, in the company of the principal, on a tour of the University of Minnesota campus) returned to school in time for the assembly and were the heroines of the next two days.

RESULTS

1. *Knowledge of the rumor and the cognitive unclarity manipulation.* Table 2 presents data on the percentage of girls who in their interviews reported that they had heard the planted rumor and the percentage of girls indicating that they knew that the principal had removed girls from their classrooms. In this and in several of the following tables, data are reported for each of the classes individually as well as for the two classes in each condition combined. Thus, the symbol CU-R1 stands for one of the two classes in the CU-R condition, CU-R2 for the other class in this condition, and so on. The symbol CU-R1 + CU-R2 stands for the two classes in the condition combined.

It is clear from Table 2 that virtually every girl had heard the planted rumor. Only one of the 96 girls interviewed reported that she had not heard the rumor. In terms just of having heard the rumor, there are no differences between conditions. Similarly, almost all of the girls, including those in the R condition, were aware that girls had been taken out of their classes. Only two girls reported that they knew nothing about this. Eighty-five per cent of the girls interviewed linked the planted rumor, in some fashion, to the manipulation.

Clearly, there had been considerable communication not only within classes but be-

tween classes. The CU girls had all heard the planted rumor. Almost all of the R girls were aware that the principal had removed girls from other classes. The unexpected volume and generality of communication make it necessary to recharacterize the three experimental conditions.

The fact that the planted rumor is widely known in both the CU-R and the CU conditions essentially reduces these two to the same condition—a situation of cognitive unclarity about an important issue, with a widely spread, planted rumor. The R condition is recharacterized as a state of cognitive unclarity about a relatively unimportant issue. Though all of the girls in this condition are aware that for some mysterious reason the principal has taken girls out of other classes, they have almost no sociometric connections with these girls. They are familiar with the faces, but indifferent to the girls. Though the manipulation poses the possibility of real changes in the relationship between the girls removed from class and their immediate classmates, no such problems exist for girls in the R condition.

Several sources of data support this recharacterization of the R condition. Previously, inferences were made about the success of the attempt to create a cognitively unclear situation from the *nature* of the questions asked the teachers. Tentative inferences relating to the importance variable may be drawn from the *number* of such questions. Virtually all subjects in all conditions knew about the cognitive unclarity manipulation. Presumably only those girls who seriously cared about the event would have attempted to glean information from the teachers. The teachers reported well over 200 questions from the 62 girls in the CU-R and CU conditions. They reported only one question from the 34 girls in the R condition.

The amount of time the girls spent speculating and discussing the manipulation may also be considered as an indication of the importance of the event. In their interviews, after they had indicated that they were aware of the manipulation and had discussed it, the girls were asked, "Could you say just how much time you've spent talking about this today? Try to make an accurate guess." Girls in the CU-R and CU conditions estimated an average of almost one hour and forty minutes, girls in the R condition estimated an average of twenty minutes. The difference is significant at better than the .001 level of confidence.[3]

Clearly these last two bits of data must be qualified by the fact that girls in the R condition necessarily learned of the events of the cognitive unclarity manipulation somewhat later than the girls in the other conditions and consequently had less time to discuss and ask questions of the teachers. However, the interviews indicate that almost all of the girls in the R condition had learned of this event by the 10:15 morning recess. Since previous to this time the girls were free to communicate to one another and to the teachers only in the two five-minute periods between classes, it is clear that this factor alone cannot account for the major differences between conditions.

Supporting evidence free of this time factor can be derived from the indications of saliency of issue in the standardized interviews with each girl. The interview was so constructed that the first question relevant to the manipulation of cognitive unclarity was a recall question and the following questions were recognition questions. After a few icebreaker questions the girl was asked: "In most schools things are pretty much the same from day to day, but sometimes things do happen that are out of the ordinary. Would you say that anything unusual happened today?" If in response to this question she failed to mention the experimental manipulation, she was asked, "Some of the other girls we've spoken to have told us that this morning some girls were

[3] It could, of course, be argued that rather than indicating that the issue was unimportant in the R condition, the number of questions to the teachers and the amount of time spent talking might indicate that there was less cognitive unclarity in the R condition. Such might indeed be the case if girls in the R condition were either (*a*) more prone to link the rumor to the manipulation and to believe that the girls taken from class had stolen the examinations or, (*b*) had some other generally accepted and credited explanation for the manipulation. The data reveal, however, that there are no between-condition differences in the extent to which the rumor is linked to the manipulation or in the degree to which it is believed that the planted rumor conclusively explains why the girls were removed from class. Further, the great majority of *S*s in the R condition indicated in their interviews that they had no real idea as to why the girls had been taken from class.

TABLE 3

MEAN NUMBER OF SUBJECTS TO WHOM COMMUNICATIONS CONCERNING THE RUMOR WERE INITIATED

Condition	*N*	Mean No. of Girls to Whom Rumor Was Transmitted
CU-R1	16	3.19
CU-R2	13	2.46
CU-R1 + CU-R2	29	2.86
CU1	18	2.22
CU2	11	2.36
CU1 + CU2	29	2.28
R1	16	1.38
R2	14	.79
R1 + R2	30	1.10

called out of class by the principal. Have you heard anything about this? What have you heard?" It is assumed that the cognitive unclarity manipulation is more salient and more important for those girls who mention it on the recall question than for those who first mention it on the recognition question. Ninety-three per cent of the girls in the CU-R condition mentioned the incident in answer to the recall question; 76 per cent of the girls in the CU condition mentioned it to the recall question; and only 26 per cent of girls in the R condition responded to the recall question by describing this event.

Such evidence supports the recharacterization of the experimental conditions. The three conditions are characterized by an equal degree of cognitive unclarity, for the event was the same mystery to all the girls. The issue involved, however, was relatively unimportant to the R condition classes and extremely important to the CU-R and the CU classes.

2. *Transmission of the planted rumor.* Our conceptualization suggests that the force to communicate a relevant rumor should vary with the degree of importance of the issue concerning which there is ambiguity. It should be anticipated then that there will be far more transmission and discussion of the rumor in the CU-R and CU conditions than in the R condition. Table 3 presents the relevant data.[4]

The data in Table 3 are derived from the responses of the girls to the interview questions concerning their knowledge of the rumor, from whom they had heard it, and to whom they had communicated it. The figures reported are the averages, for each class and condition, of the number of different girls to whom each girl is reported as *initiating* a communication concerning the rumor. There is an average of more than twice as many transmissions of the rumor in both the CU-R or CU conditions than in the R condition. The difference between CU-R and R conditions is significant at better than the .01 level of confidence; between CU and R conditions at the .03 level of confidence. There is no significant difference between CU-R and CU conditions.

Transmission of the rumor was widespread in CU-R and CU classes and relatively restricted in the R condition classes. Seventy-eight per cent of the girls in CU-R and CU conditions initiated one or more communications concerning the rumor; only 40 per cent of the girls in the R condition did so. Since we know that virtually all of the girls heard the rumor, it would seem that knowledge of the rumor creates far stronger forces to communicate and discuss it when the issue to which it is relevant is important than when it is unimportant. Some caution, however, must be observed in weighing this interpretation. It was previously suggested that the dramatic nature of the manipulation probably produced surprise and excitement as well as a state of cognitive unclarity. It would seem a reasonable guess that such factors would operate more strongly for the girls who were immediate witnesses of the event than for those girls who heard about the incident but were not present. The CU-R and CU conditions, then, may differ from the R condition not only in the importance of the event, but in the effects produced by being witness to the manipulation. Though we do believe that importance is the crucial variable in accounting for the differences between conditions, it is clear that, within the present design, this conclusion must be tempered by the existence of these additional factors.

[4] In order to make the data of all six classes completely comparable, the data for the girls with whom the rumor was originally planted are not included in this table. In the CU-R condition, these four girls initiated communications concerning the rumor with an average of 7.00 girls; in the R condition, they averaged 6.00 girls.

It is to be expected, too, that the importance of the issue varied within class and condition. For some girls the issue may have been extremely important and for others trivial. One criterion for distinguishing among the girls on this dimension would be the nature of their relationship to the girls who are taken out of class by the principal. Plausibly, this should be a more important event for those who are good friends of the girl removed from class than for those who are not, and the friends should be expected to communicate the rumor more.

Since almost all sociometric choices were within class, the following means are based on just the four classes in the CU-R and CU conditions. Girls who on their sociometric have given one of their choices to the girl taken out of their class by the principal initiated communications concerning the rumor to an average of 3.10 different girls. Girls who did not make this sociometric choice initiated only 2.00 communications of the rumor. This difference is significant by t test at the .08 level of confidence.

3. *Origin of new rumors.* The same factors which promote the spread of the planted rumor should presumably stimulate the kind of speculation and guesswork from which new rumors arise and it should be anticipated that there will be greater diversity and circulation of new rumors in those conditions where the issue is most important. Relevant evidence is presented in Table 4.

In column 3 of Table 4 are reported the percentages of girls in each of the classes who in their interviews reported that they discussed some rumor related to the cognitive unclarity manipulation other than the planted rumor. Some 70 per cent of the girls in the CU-R and CU conditions reported discussing other rumors; less than 15 per cent of the girls in the R condition did so. The difference between either CU-R or CU condition and R condition is significant at better than the .001 level of confidence. The difference between CU-R and CU conditions is not significant. Clearly the greater the importance of the issue, the greater the circulation of new rumors.

Not only is there greater circulation of rumors other than the planted rumor in the two high importance conditions but there is greater diversity of new rumors as well. In column 4 of Table 4 are listed the number of different rumors in circulation in each of the classes. There is an average of over 12 different rumors per class reported in the CU-R and CU conditions and of 1.5 different rumors per class in the two R condition classes.

In terms of our previous reasoning it might be anticipated that "good friends" would tend to circulate new rumors more than would those girls who do not choose sociometrically the girls taken out of class by the principal. Though friends circulated an average of 1.55 new rumors and nonfriends an average of 1.32, a slightly smaller proportion of friends transmitted such new rumors. Sixty-eight per cent of friends reported discussing rumors other than the planted rumor and 74 per cent of the nonfriends did so. Neither of these differences is significant.

There is, however, a difference between these two groups in the kind of rumor which they report having discussed. It is possible to categorize the various rumors circulated in terms of the optimism or pessimism of the explanation offered for the removal of the girl from class. Some of these rumors were distinctly favorable to this girl, e.g., "She's a great beauty and has been invited to tea at the principal's house." Other rumors were more unfavorable, "She's being disciplined for going to a wild party last weekend." Still other rumors were neutral in tone, "She's going to attend a lecture." Table 5 presents the percentage of each type of rumor transmitted by friends and nonfriends. Of the

TABLE 4

THE PREVALENCE AND VARIETY OF NEW RUMORS

Condition	N	Percentage of Girls Reporting New Rumors	Number of Different Rumors
CU-R1	18	72.2	16
CU-R2	15	80.0	15
CU-R1 + CU-R2	33	75.8	
CU1	18	72.2	14
CU2	11	54.6	5
CU1 + CU2	29	65.5	
R1	18	5.6	1
R2	16	25.0	2
R1 + R2	34	14.7	

TABLE 5
FRIENDSHIP AND NEW RUMORS

Group	N	Number of Times New Rumors Transmitted	Percentage of New Rumors Which Were:		
			Favorable	Neutral	Unfavorable
Friends	31	48	52	19	29
Nonfriends	31	41	34	10	56

total number of rumors transmitted by friends 52 per cent are favorable and 29 per cent unfavorable. Nonfriends transmitted 34 per cent favorable rumors and 56 per cent unfavorable. Of the friends who did transmit rumors 76.2 per cent transmitted one or more favorable rumors. Of nonfriends who transmitted rumors, 47.8 per cent transmitted one or more favorable rumors. This difference is significant at close to the .10 level of confidence. Though there is no difference between friends and nonfriends in the extent to which they transmit rumors other than the planted rumor, there is a tendency for friends to transmit predominantly favorable rumors and for nonfriends to transmit predominantly unfavorable rumors.

Distortion of the planted rumor. The interview of each girl was examined for indications of distortion of the planted rumor. In no one of the 96 interviews is there any indication of distortion. In every case, the planted rumor was reported to the interviewer in essentially the form in which it was originally planted with no instance of embellishment or variation. Though many new and bizarre rumors did spring up, the planted rumor itself came through a day's discussion intact. This finding is consistent with the results of two of the studies (2, 6) discussed in the introduction to this paper where also there was no indication of distortion of the planted rumors. Caplow, too, in a study of rumors in war (4) reports "the veracity of rumors did not decline noticeably during transmittal." Though he notes the existence of tendencies to distortion, his major impression is that of a marked lack of distortion.

Such findings, of course, are in direct contrast to the widespread impression of rumor as an unreliable and sometimes fantastically distorted form of communication. These results suggest that experiments such as those of Bartlett (3) and Allport and Postman (1) are questionable laboratory paradigms of the field situation of rumor transmission. This is not meant, in any way, to question the results of these laboratory and classroom experiments as such, but to suggest that the results of these studies are not immediately applicable to field situations in which rumor transmission is a voluntarily initiated action subject to a variety of corrective tendencies, rather than an experimentally induced communication with no possibility of correcting exaggerations, distortions, or omissions during the course of transmission.

Caplow in discussing the discrepancy between his findings and those reported in the Allport and Postman experiments suggests two mechanisms by which distortion may be eliminated or prevented during the course of transmission of a rumor. (*a*) "A rumor is usually heard more than once, and usually transmitted more than once by each individual in the channel. This re-circulation tends to eliminate variation. . . ." (*b*) Persons associated with previous inaccuracies or exaggerations tend to be excluded from the developing channels of rumor transmission (4, p. 301).

Other factors which may account for the differences between the field studies discussed and the Allport-Postman experiments are: (*a*) The complexity of the material transmitted. In the laboratory experiments, the material transmitted is customarily the content of highly detailed and complicated pictures or stories. In the field studies the planted rumors have usually been relatively simple, uncomplicated, and with a minimum of irrelevant detail. Plausibly, the process of leveling (the omission of details in serial reproduction) will be most apparent when highly detailed and complicated material is transmitted.

(*b*) The nature of the force to communicate. In the laboratory experiments, subjects communicate the contents of picture or story only to follow the experimenter's instructions. In the field studies, subjects communicate only on their own initiative. Presumably, subjects will transmit a rumor on their own initiative only when the content of the rumor is of some interest to them. Higham (7) in a study using the technique of serial reproduction has main-

tained that there is less distortion when subjects are ego-involved and interested in the content of their communication.

SUMMARY

Current notions of the determinants of rumor spread and distortion were tested in a field experiment. It has been hypothesized that rumors will spread when there is (*a*) a state of cognitive unclarity about (*b*) an important issue which is (*c*) common to all or most members of a group. Evidence is presented indicating that under conditions of widespread cognitive unclarity there is far more transmission of a planted rumor and far more speculation involving new rumors when the issue is important than when it is relatively unimportant.

In distinct contrast to expectations created by studies using the technique of serial reproduction, there is absolutely no indication of distortion of the planted rumor. Several factors are suggested which may account for this difference between laboratory and field situations.

REFERENCES

1. ALLPORT, G. W., & POSTMAN, L. *The psychology of rumor*. New York: Holt, 1947.
2. BACK, K., FESTINGER, L., HYMOVITCH, B., KELLEY, H. H., SCHACHTER, S., & THIBAUT, J. W. The methodology of studying rumor transmission. *Hum. Relat.*, 1950, **3**, 307–317.
3. BARTLETT, F. C. *Remembering*. London: Cambridge Univer. Press, 1932.
4. CAPLOW, T. Rumors in war. *Soc. Forces*, 1947, **25**, 298–302.
5. FESTINGER, L., CARTWRIGHT, D., BARBER, K., FLEISCHL, J., GOTTSDANKER, J., KAYSEN, A., & LEAVITT, G. A study of a rumor: its origin and spread. *Hum. Relat.*, 1948, **1**, 464–485.
6. FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures of informal groups*. New York: Harper, 1950.
7. HIGHAM, T. M. The experimental study of the transmission of rumor. *Brit. J. Psychol.*, 1951, **42**, 42–55.
8. SCHALL, H. M., LEVY, B., & TRESSELT, M. E. A sociometric approach to rumor. *J. soc. Psychol.*, 1950, **31**, 121–129.

Received March 8, 1954.

CHANGES IN GOAL-OBJECT PREFERENCE AS A FUNCTION OF EXPECTANCY FOR SOCIAL REINFORCEMENT[1]

DAVID E. HUNT

Yale University

THE effectiveness of social rewards in the modification of an individual's goal preferences has been frequently observed. This effect is perhaps most clearly illustrated by the development of a young child's goal preferences in relation to the reaction of other individuals in his environment.

In studying the conditions which lead an individual to modify his goals, previous investigators have shown that certain factors are related to changes in preference for both goal objects and for activities. Gebhard (5) found that preferences for an activity increased after an experience of success with that activity, i.e., an experimenter (*E*) informed a subject (*S*) that he had done very well, citing false norms to support this statement; preferences decreased conversely after failure. A subsequent study by Gebhard (6) indicated that the expectation of success or failure verbally induced by *E* before the activity was a significant factor in accounting for the resulting preference changes. Studying goal-object preferences, Filer (4) found that preferences increased for a goal object after attainment and decreased after nonattainment. Duncker (2) found social suggestion (either in the form of adult verbalization or through observation of peer behavior) to be effective in modifying (food) goal preferences.

The present investigation was designed to study one aspect of the relationship between social reinforcement and goal-preference changes within the framework of a social learning theory of personality developed by Rotter and his students (8, 9). One of the two central constructs in this theoretical position[2] is that of reinforcement value, which is defined as "The degree of preference for any reinforcement to occur if the possibilities of their occurring were all equal" (9, p. 112). This value may be either positive or negative.

From a theoretical point of view, reinforcement value is probably most similar to Lewin's valence. At an operational level, this construct is quite similar to the more generic term, goal-object preference, and this latter term will be used in the remainder of this paper as equivalent to reinforcement value.

According to this social learning point of view (8), changes in goal-object preference are considered to be dependent upon two factors: (*a*) the *preference value* of the reinforcement which follows activity with that goal object and is related to it by the individual, and (*b*) the *expectancy* that this reinforcement will occur. The effect of the value of the subsequent reinforcement on preference change may be better understood through an example. If a child's activity with toy A elicited only a noncommittal grunt from his parent and the child's activity with toy B was greeted by enthusiastic verbal approval, it would be predicted that the child's preference would increase more for toy B than for toy A.

Dunlap (3) has experimentally demonstrated this sort of relationship between changes in goal-object preference and the value of subsequent reinforcement. He measured change in preference by asking each *S* to rank nine toys in order of ownership preference before and after an experimental period. During this experimental period, each *S* carried out a standard prescribed task with a critical toy and the *E* administered one of the four experimental verbal reinforcements in relation to this activity. The preference changes for the two positive conditions (strong approval and approval-plus-expectancy for extrasituational approval) were significantly

[1] Adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Psychology, Ohio State University. The writer wishes to express his appreciation to Professors Julian B. Rotter, George A. Kelly, and Delos D. Wickens for their helpful suggestions and advice. The writer is indebted to Russell Slater, director of the Child Study Division, and to the principals and teachers of the participating schools in the Columbus Public School System for their cooperation in making subjects available. A summary report of this study was given at the APA meeting, Cleveland, September, 1953.

[2] For a more detailed presentation of this point of view, see Rotter (8).

greater than the changes for the neutral and mild negative conditions.

The present study, a companion experiment to Dunlap's, is concerned with the second factor, the *expectancy* for the subsequent reinforcement to occur. In a social learning context, expectancy is defined as "the subjective probability held by the individual that a particular reinforcement will occur as a function of, or in relation to, a specific behavior in a given situation or situations" (9, p. 112). The operation for expectancy used in the present study was based on the previous occurrence or nonoccurrence of a reinforcement. It should be noted that this use of controlled previous trials is only one of several possible operations for estimating expectancy.

An example may again be in order: if a child's activity with toy A is only *occasionally* followed by verbal approval from an adult, but activity with toy B is *always* followed by verbal approval, it would be predicted that the child's preference would increase more for toy B than for toy A. It is important to note that the verbal statements referred to in this example are assumed to be of approximately equal value.

If the preference for a goal object is determined by the two factors of value of the subsequent reinforcement and the expectancy for its occurrence, it should be possible to hold one of these factors constant in order to study the effect of the other on changes in preference. For example, in the Dunlap study (3), the factor of expectancy was presumably controlled since the subsequent reinforcement occurred only once in each of the four groups.

In the present investigation an attempt was made to hold the value of the subsequent reinforcement constant, varying the expectancy for its occurrence. Therefore, it was hypothesized that changes in goal-object preference would be directly related to the expectancy for subsequent reinforcement. The specific hypotheses which were formulated from this general hypothesis will be presented in the following section after a description of the experimental design.

METHOD AND HYPOTHESES

Materials

Ten goal objects were chosen: an Erector set, a Sand Pen art set, a box of 26 game materials, a pocket knife, a three-bladed hand saw, a set of American Plastic Bricks, a set of Lincoln Logs, a Senior Tinkertoy set, a ring toss game, and a set of jigsaw puzzle pictures of Indian scenes. This group of toys was, with minor exceptions, the same as the group used by Dunlap (3) which he had previously selected by eliminating those toys ranked either consistently high or low. The writer further selected four of these toys—the set of American Plastic Bricks, the set of Lincoln Logs, the Senior Tinkertoy set, and the ring toss game—to be used as critical goal objects in the reinforcement phase. Selection of these four toys was based on preliminary work which indicated a tendency for them to be ranked in the middle range of the ten objects.

TABLE 1

SCHEDULE OF REINFORCEMENT CONDITIONS

Condition	Monday	Tuesday	Wednesday	Thursday	Estimated Expectancy	Frequency
I	+	+	+	+	1.00	4
II	No activity	No activity	+	+	1.00	2
III	0	+	0	+	.50	2
IV	0	0	0	0	0	0

Note.—The + refers to occurrence of experimental reinforcement; 0 refers to nonoccurrence of experimental reinforcement.

Experimental Design

The experimental reinforcement consisted of praise from *E* in relation to *S*'s carrying out a prescribed task with one of the four critical goal objects. This experimental reinforcement occurred according to the pattern in Table 1 so that the conditions differed with regard to expectancy for reinforcement, i.e., its occurrence or nonoccurrence. An intra-individual design was employed so that each *S* served in each of the four reinforcement conditions, a different toy being assigned to each condition.

Since each *S* expressed a rank-order preference for all ten goal objects both before and after the reinforcement phase, changes in these rankings provided a measure of the change in preference attributable to the reinforcement condition to which the goal object had been assigned. The four critical objects were assigned in a counterbalanced fashion so that each one appeared an equal number of times, i.e., ten, in each of the four conditions.

Specific Hypotheses

It will be seen from Table 1 that the experimental design was devised so that the expectancy was varied between the conditions, thus permitting differential predictions. From the central prediction that changes in goal-object preference will be directly related to expectancy for subsequent reinforcement, the following hypotheses, stated in experimental form, were formulated:

Hypothesis 1. That there will be greater positive preference change in condition I than in condition III.

Hypothesis 2. That there will be greater positive preference change in condition III than in condition IV.

An inspection of Table 1 reveals that the preceding predictions, if sustained, may be accounted for by an interpretation based either on expectancy or on the frequency of occurrence. That is, conditions I, III, and IV occupy the same relative position with regard to expectancy (1.00, .50, and 0) and frequency (4, 2, and 0). Therefore, condition II was instituted, equal to condition III in frequency. Since these two conditions presumably differed only in terms of expectancy, the following hypothesis was formulated:

Hypothesis 3. That there will be greater positive preference change in condition II than in condition III.

Subjects and Procedure

The experimental procedure to be described extended over a two-week period. Fourth-grade boys from four public elementary schools, ranging from eight to eleven years of age, served as *S*s. Preference change data to test the hypotheses were based only on the 40 *S*s who completed all phases of the experiment.

Preranking phase. The 10 goal objects were displayed on a large table. Each *S* came to the experimental room individually. He received the following instructions: "I am interested in knowing more about what toys boys like best and I would like for you to help me out. As you can see, there are a lot of toys here (indicating display). These are not to give away, but I would like to show them to you. Let's begin by taking a look at each one so that you'll know what they are all like."

After *E* pointed out and gave a brief description of each toy, he continued: "Now that you have seen all of the toys I would like for you to do something. Look them all over carefully. Then tell me, if you could have any one of these toys you wanted,[3] which one would you most like to have for your own?"

As each toy was selected, *E* recorded the choice, removed the toy, and repeated the question. The *S*'s expressed rank for a goal object was taken as a measure of its initial preference value.

Every boy in each class participated in this phase of the experiment. However, any *S* who ranked any one of the four critical goal objects either first or last was not carried further in the experiment since such an initial ranking precluded the possibility for that object to change in one direction (this criterion accounted for the elimination of 76 *S*s).

This preranking phase was carried out on Friday. For every *S* who continued on to the reinforcement phase[4] one of the four critical toys was then assigned in a counterbalanced fashion to one of the reinforcement conditions.

Reinforcement phase. The critical goal objects (three on the first two days and four on the last two days) were placed on separate small tables and covered by a cloth. The instructions for this phase were: "Today we are going to look at some of the toys which you saw last week and you will have a chance to play with them. We won't be looking at all of the toys but we will look at several of them. Let's begin with this one." The *E* then directed *S* to one of the tables, removed covering cloth and said:

(For construction toys) "Here is a (name of task) which has been made with these (name of toy). Here's the way it goes (brief demonstration of assembling the model). I would like for you to look this over and then make one like it using the materials in this box."

(For ring toss) "I would like for you to throw some rings at this stake which I will put over here. You can stand here (indicating line on floor) and throw four rings and then move to another place and throw the rings."

The prescribed tasks were the same for all *S*s. The *E* reacted to *S*'s activity in terms of its position in the reinforcement schedule in Table 1.

For *occurrence of reinforcement* *E* interjected the following comments at appropriate intervals with an attempt to make each statement once: "That was very good." "You do very well with these (name of toy)." "You caught on to the idea right away." "I like the way that you (specific behavior mentioned)."

For *nonoccurrence of reinforcement*, after giving the task instructions, *E* said (for ring toss) "You go ahead and throw rings at the stake" or (for construction toys) "you go ahead and make one like this and let me know when you have finished." The *E* then went to his desk and occupied himself either for three minutes (for ring toss) or until *S* indicated that he had completed the task (for construction toys).

For example, on Monday there was no activity with the condition II toy, *E* reacted positively to *S*'s activity with condition I toy, but did not react to activity with toys assigned to conditions III or IV. Order of presentation was counterbalanced. Ten to twenty minutes were required for *S* to carry out the three (or four) experimental tasks.

The *S* continued to carry out different tasks with the goal objects on the remaining three days while *E* reacted according to the schedule in Table 1.

Postranking phase. On Friday, the day following completion of the reinforcement phase, all ten goal objects were again displayed, and *E* gave the following instructions: "I would like for you to look at the toys again today. Look them over carefully. Then tell me, if you could have any one of these toys that you wanted, which one would you most like to have for your own? This may be the same as before, or it may be different —that's up to you. I would like to know which one you would most like to have *today* if you could have any one of them for your own."

This reranking procedure was repeated again eight days after the completion of the reinforcement phase. At the conclusion of this final ranking, *S* was asked if he owned any of the toys.

RESULTS

Analysis of data was based on changes in ranking for each of the critical goal objects

[3] This qualifying phrase was used in the instructions in order that the ranking question satisfy the definition of reinforcement value, "degree of preference for any reinforcement to occur *if the possibilities of their occurrence were all equal.*"

[4] Of the 48 *S*s retained, eight were eliminated on the basis of two predetermined criteria: absence at any one of the experimental sessions (six *S*s) or inability to complete one of the experimental tasks (two *S*s).

assigned to a particular condition. The test of the hypotheses, therefore, was based on comparisons of the rank-difference-score means for the conditions in question. This procedure appeared to be justified since the four conditions were counterbalanced with respect to toys and were, as will be noted in Table 2, almost identical with regard to initial mean rank.

One-day preference changes. Before proceeding to an analysis of variance, Bartlett's test of homogeneity of variance was applied and the assumption of homogeneity was supported. Testing for the normality of the distribution by goodness of fit also indicated that the one-day change data met the requisite assumption.

Table 3 indicates that only the difference attributable to reinforcement conditions was significant. Neither differences attributable to toys alone nor to interaction between toys and conditions were significant.

In order to test the hypotheses under consideration, mean change scores for the reinforcement conditions were compared. The results summarized in Table 4 indicate that Hypothesis 1 (I > III) and Hypothesis 2 (III > IV) are supported, but Hypothesis 3 (II > III) is not supported. In fact, the direction of the insignificant difference in Hypothesis 3 is opposite to the direction predicted.

Eight-day preference changes. Although the eight-day change scores met the assumption of homogeneity of variance, the distribution was markedly leptokurtic. Regarding this, Cochran states "The consensus . . . is that no serious error is introduced by non-normality in the significance levels of the F-test or of the two-tailed t-test" (1, p. 24). Therefore an analysis of variance was again carried out.

Table 5 indicates that the reinforcement conditions continued to provide the only significant source of variance, although the interaction approached significance ($.10 > p > .05$). When we compared mean change scores by the two-tailed t test, none of the three hypotheses were supported, although the four conditions remained in the same relative position as found on the one-day rankings. It may be noted, however, that condition I was significantly ($p < .01$) greater than condition IV. Although the differences between conditions were less pronounced, the total mean increase for all four conditions taken together remained the same as on the first reranking. This finding might be interpreted as indicating that the generalized effect of the social reinforcement was maintained over time, but that the specific effects were diminished.

Effect of other factors. The *S*s reported 18 instances of ownership of the experimental

TABLE 2

MEAN SCORES AND MEAN RANK CHANGES

Condition	Prerank Mean	One-Day Mean	One-Day Change	Eight-Day Mean	Eight-Day Change
I	5.90	4.25	+1.65	4.55	+1.35
II	5.85	5.52	+.33	5.32	+.53
III	5.72	5.22	+.50	5.12	+.60
IV	5.85	6.20	−.35	6.05	−.20

TABLE 3

ANALYSIS OF VARIANCE OF THE EFFECT OF REINFORCEMENT CONDITION AND KIND OF TOY ON ONE-DAY CHANGE SCORES

Source	*df*	Mean Square	*F*	*p*
Conditions	3	27.62	5.75	<.001
Toys	3	10.46	2.18	
Conditions × Toys	9	4.39		
Within conditions	144	4.81		
Total	159			

TABLE 4

DIFFERENCES BETWEEN ONE-DAY MEAN CHANGE SCORES OF REINFORCEMENT CONDITIONS

Hypothesis	Prediction	Difference	*t*	*p**
1	I > III	1.15	2.35	<.02
2	III > IV	.85	1.73	<.05
3	II > III	−.17		

Note.—$df = 78$; $SE_m = .49$.

* One-tailed test of the distribution of t was used.

TABLE 5

ANALYSIS OF VARIANCE OF THE EFFECT OF REINFORCEMENT CONDITION AND KIND OF TOY ON EIGHT-DAY CHANGE SCORES

Source	*df*	Mean Square	*F*	*p*
Conditions	3	15.83	3.08	<.05
Toys	3	2.42		
Conditions × Toys	9	9.86	1.92	
Within conditions	144	5.12		
Total	159			

toys out of 160 possible instances (40 *S*s × 4 toys). Changes associated with these owned toys did not differ appreciably from those changes occurring in the nonowned toys, however, so that the effect of ownership was apparently negligible.

Other factors such as age, intelligence, desire for adult approval, etc. were controlled by the intra-individual design.

DISCUSSION

The results reported in support of Hypotheses 1 and 2 indicate a relationship between increased preference for a goal object and either the expectancy for, or the frequency of, subsequent social reinforcement. As mentioned earlier, however, these two hypotheses do not permit a distinction between interpretations of expectancy or frequency.

Hypothesis 3, which was intended to provide such a distinction, received no support. Before concluding that the lack of evidence in support of Hypothesis 3 necessitates accepting a frequency formulation as the most useful interpretation, it may be worth while to consider the two conditions involved in this hypothesis in more detail.

The unsupported prediction rested on the assumption that the only meaningful difference between condition II and condition III was in expectancy. Are there any other differences between these conditions which might have accounted for the lack of differences? Two possibilities were considered. The first was that the activity with the toy itself, i.e., Monday and Wednesday periods in condition III, might have been positively reinforcing and thus have masked differences between the conditions. Since condition IV was a "control" condition with activity but no experimental reinforcement, preference changes in this condition were recomputed in relation to the six noncritical goal objects.[5] No appreciable increase was noted, however, so that the effect of the activity per se as reinforcing must be regarded as unlikely.

A second possibility considered was the difference in serial position of the reinforcements in the two conditions, i.e., the reinforcements in condition II occurred on the last two experimental days. If *S*'s experience over the four-day period is regarded as a unit, the first reinforcement for condition II occurred after reinforcements had already occurred in conditions I and III. Jessor (7) found that successive statements of verbal approval become decreasingly effective, or stated in other terms, that the value of each successive reinforcement in a series is reduced according to its later position in that series. Such a reducing effect may have operated against the predicted results in the present study since the reinforcements in condition II occurred late in the series.

Although this reducing increment hypothesis must be regarded as speculative with regard to the present data, it could be tested by employing two *groups* of *S*s. Each group would experience reinforcement on the same day and the "frequency" group would experience intervening trials of no reinforcement while the "expectancy" group would have no experience between the reinforced trials.

Nature of activity. The importance of considering the nature of the task in assessing the effect of incentives has been previously emphasized by Wolf (10, p. 39). In this respect, it is noteworthy that neither the kind of toy nor the interaction between toys and conditions contributed significantly to the one-day preference changes. In other words, the differential effects of reinforcement were apparently relatively independent of the nature of the task.

SUMMARY

The present study was designed to investigate the relationship between changes in goal-object preference and expectancy for social reinforcement following activity with a goal object.

Preference measures consisted of *S*'s ranking ten preselected toys. An intra-individual design was employed in which each of 40 fourth-grade boys served in four reinforcement conditions. The *S* performed tasks with four critical toys over a four-day experimental period. The *E*'s verbal approval to *S*'s activity was considered the experimental reinforcement, and the expectancy for reinforcement

[5] Necessitated because in the intra-individual design, the rankings of the four critical objects are relative to each other as well as to the other six objects; it is therefore possible that an increase in ranking for a toy in one condition might necessarily lead to a decrease in value for a toy in another condition.

was varied by the occurrence or nonoccurrence of this reinforcement.

Results of the study may be summarized as follows:

1. Following a constant number of experiences with several goal objects, the resulting preference changes measured after one day appeared to be related to either the expectancy for, or the frequency of, social reinforcement.

2. This relationship appeared to be relatively independent of the nature of the task.

3. Changes in preference value after eight days were less specific to reinforcement conditions so that the differences between conditions were not significant.

4. The present results supported a frequency of reinforcement interpretation of the preference changes which occurred. The lack of support for an expectancy formulation was considered in relation to the present design; the possibility of clarifying the relative efficacy of these two interpretations through a different experimental design was discussed.

REFERENCES

1. COCHRAN, W. G. Some consequences when the assumptions of analysis of variance are not satisfied. *Biometrics*, 1947, **3**, 22–38.

2. DUNCKER, K. Experimental modification of children's food preferences through social suggestion. *J. abnorm. soc. Psychol.*, 1938, **33**, 489–507.

3. DUNLAP, R. L. Some effects of social reinforcement on changes in attractiveness of children's goal objects. *Amer. Psychologist*, 1953, **8**, 344. (Abstract)

4. FILER, R. J. Frustration, satisfaction, and other factors affecting the attractiveness of goal objects. *J. abnorm. soc. Psychol.*, 1952, **47**, 203–212.

5. GEBHARD, MILDRED E. The effect of success and failure upon the attractiveness of activities as a function of experience, expectation, and need. *J. exp. Psychol.*, 1948, **38**, 371–388.

6. GEBHARD, MILDRED E. Changes in the attractiveness of activities: the effect of expectation preceding performance. *J. exp. Psychol.*, 1949, **39**, 404–413.

7. JESSOR, SHIRLEY L. The effects of reinforcement and of distribution of practice on psychological satiation. Unpublished doctor's dissertation, Ohio State Univer., 1951.

8. ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

9. ROTTER, J. B., FITZGERALD, B. J., & JOYCE, J. N. A comparison of some objective measures of expectancy. *J. abnorm. soc. Psychol.*, 1954, **49**, 111–114.

10. WOLF, THETA H. *The effect of praise and competition on the persisting behavior of kindergarten children*. Minneapolis: Univer. of Minnesota Press, 1938.

Received March 19, 1954.

LEARNING AND RETENTION OF WORDS OF SEXUAL AND NONSEXUAL MEANING[1]

MAURICE G. KOTT

Department of Institutions and Agencies, State of New Jersey

CLINICAL psychologists and psychiatrists have gathered clinical evidence for the Freudian theory of repression. Laboratory studies, however, have failed to demonstrate or reject conclusively the existence of memory defect or forgetting as means of enduring a motivational conflict. The many studies of recall of pleasant and unpleasant materials, of word association results, and of the effect of interruption in the manner of the Zeigarnik experiment have been thoroughly reviewed in the past. Zeller (13) maintains that while differential forgetting of unpleasant experiences has been demonstrated, such findings could be interpreted as the results of differences in learning, motivation, and set rather than repression. Rapaport (6) comments that the evidence in support of repression is slight, although the classic association theory is insufficient to account for the demonstrated results. He criticizes the various experimental studies because of the multiplicity of factors called emotional. Sears (7) is of the opinion that such studies fail to demonstrate clearly the existence of repression. He calls particular attention to the fact that Freud's concept of repression deals with impulses which are instinctual, or at least conflicting with conscience, and not with ideas judged to be unpleasant on any other grounds.

With reference to this comment by Sears, it is worth noting that Freud states, "Theoretically there is no objection to supposing that any sort of instinctual demand whatever could occasion these same repressions and their consequences; but our observation shows us invariably so far as we can judge, that the excitations that play this pathogenic part arise from the component instincts of sexual life" (3, p. 85). Consequently, experimentation on the theory of repression should touch upon sexuality, which Freud conceived as having an invariable role in the process.

Freud recognized two kinds of repression, "primal" and "proper." "Primal repression" follows the forcible thwarting of an instinctual activity with the result that any ideational or behavioral form of the instinct is kept from consciousness. It is a means of avoiding anticipated punishment for behavior which is related to the original instinctual activity that was a cause of pain (1). This kind of repression is not at present a feasible area for experimental study for obvious technical and ethical reasons. "Repression proper" concerns "mental derivatives of the repressed instinct-presentation, or such trains of thought as, originating elsewhere, have come into associative connection with it. On account of this association these ideas experience the same fate as that which underwent primal repression" (1, p. 86). Moreover, Freud states elsewhere, "Repression denies to the rejected idea . . . (its) translation . . . into words" (2, p. 134). This implies that words which relate directly to the sexual instinct will be subjected to "repression proper."

It therefore appears that the concept of repression proper would lead to the expectation that words of sexual meaning would be learned and retained with less efficiency than words of nonsexual meaning. However, direct response with sexual words may be a source of embarrassment to subjects (*S*s) and difference in performance with sexual and nonsexual words could be a function of conscious avoidance (suppression) as well as repression. It is necessary to enable *S* to respond without direct embarrassment. Such control could be established by offering the sexual and nonsexual words as material to be learned in paired association with "neutral" symbols of some sort (for example, nonsense syllables). The paired-association technique appears to be allowable within the framework of the general theory of repression, since Freud has held that forms of ideation behaviorally unrelated to the

[1] This paper is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to the Graduate School of the University of Pennsylvania in January, 1953. Appreciation is expressed to members of the writer's committee, Drs. F. W. Irwin, M. Murphy, and J. Wischner, and to the staffs of the New Jersey State Hospitals at Trenton and Greystone.

instinct but associated to the repressed idea display the same sensitivity as the idea originally repressed. Excitation in these areas gives rise to anxiety, which is then the basis for fresh repression to avoid further development of anxiety (2, p. 116). It is therefore expected that associates to sexual words will be repressed as sexual words themselves might have been.

However, it is possible that the lesser efficiency anticipated in the learning and retention of associates to sexual than associates to nonsexual words might result from the shock value which many body or anatomical words possess as well as the specific meaning of sexual words. Consequently, the nonsexual words referred to previously should be of two varieties: "neutral" words that have no reference to sex or body, and words which do not refer to normal sexual structure or function but are related to human physiology or anatomy and may be considered unpleasant in the sense that they are likely to be avoided in ordinary conversation.

The preceding discussion leads to the formation of the following hypothesis:

Hypothesis I. Associates to sexual words will be learned and retained less efficiently than associates to neutral or unpleasant words.

A further important characteristic of Freud's concept of repression is its relation to anxiety. An early view of Freud's was that the energy of the instinct detached from the idea by repression continues as a "charge of affect" (1, p. 87; 4). It was "the immediate fate of the affect to be converted into anxiety" (4, p. 350; 3). However, a later theory considers repression as a device for handling the anxiety which occurs when certain instinctual impulses are activated (5). Hence, it is to be expected that anxious *Ss* when confronted with sexual words will exhibit repression to a greater degree than nonanxious *Ss*. This may be expressed in the following hypothesis:

Hypothesis II. Anxious *Ss* will show a greater relative difficulty in learning and retention of associates to sexual words as compared with associates to neutral and unpleasant words than will normal *Ss*.

METHOD

Experimental design. Three lists were made up—a list of sexual words, a list of unpleasant words, and a list of neutral words. A reasonably close matching among lists with regard to number of noun and verb forms, mean number of letters, vowels, and consonants was attained. The frequency of appearance of these words in a semantic count is shown in Table 1.

TABLE 1

FREQUENCY OF APPEARANCE IN SEMANTIC COUNT (THORNDIKE-LORGE WORD LIST)

Frequency	Word List		
	Sex	Unpleasant	Neutral
Appearing at least once per million words	6	5	7
Appearing between once per million and once per four million words	2	4	2
Appearing 4 times per 18 million words	0	0	0
Not appearing	1		

Nine four-letter nonsense syllables were selected from Woodworth's list (12). These were systematically varied in their assignment to the words of a list so that nine sets of paired associates for each list were developed. The lists and nonsense syllables follow: *Sex*, douche, prostitute, seduce, sperm, penis, erection, rape, vagina, intercourse; *Unpleasant*, rectum, expectorate, bowels, scab, diarrhea, hernia, vomit, urine, excretion; *Neutral*, dumpling, bounty, culture, opiate, reaction, appear, twister, flotsam, incline; *Nonsense syllables*, jish, glet, crad, lerm, sark, fape, kise, roif, twic.

Subjects. One hundred and twenty males participated in the experiment. Sixty normal *Ss* were assigned at random to three word-list groups of 20 each; the same was done with 60 anxious *Ss*. The *Ss* were between 20 and 40 years old; according to Wechsler (11) this limits the anticipated intellectual deterioration to 4 per cent. They had IQ's between 85 and 115 as represented by scores on the CAS short form of the Wechsler (9). All had achieved at least the eighth grade, and none had completed a year of formal education beyond high school. The *Ss*, and also their fathers, were in the semiprofessional and business, skilled, or semiskilled labor groups as defined by Terman (10) in his revision of the Taussig five-grade classification. None had participated in a psychological experiment previously and all were informed, "The purpose of this experiment is to determine speed of learning and it will involve some words which deal with sex and sometimes are called dirty." All *Ss* willingly participated in the task.

The means and standard deviations of age, CAS scores, and education distributions show that as far as these factors are concerned, the groups were well matched and there were no significant differences between the normal and anxious groups or among the various conditions of the experiment.

The normal group was composed of attendants at the New Jersey State Hospital at Trenton. They were screened by a questionnaire of seven questions derived in part from Singer (8) and resembling the Cornell Selectee Index. It is recognized that a group drawn from a single occupational source does not constitute

a random sample from the general population of normal persons.

The anxious group was composed of men who were legally committed to and resident at the New Jersey State Hospitals at Trenton and Greystone Park. A diagnosis of neurosis, or schizophrenia with anxiety, had been established on all through staff conference. Most diagnoses were of neurosis; however, three *Ss* learning the sex list, three learning the unpleasant list, and five learning the neutral list were diagnosed as "schizophrenic with anxiety." None were undergoing, or had undergone in the past four weeks, any surgical or convulsive treatment or other therapy which might result in impairment of memory function. All were capable of establishing rapport with *E*. The schizophrenic patients were included because of the unavailability of a sufficient number of neurotics. It was recognized that hospitalized patients, especially those diagnosed as schizophrenics, may have been more seriously disturbed than the patients on the basis of whom Freud formulated his concept of repression. However, their use was felt to be warranted because of the presence of anxiety which was critical for testing Hypothesis II. (Inspection of the results failed to disclose anything atypical in the scores of the schizophrenics as compared with the other anxious *Ss*.)

Procedure. The method of adjusted learning was utilized and the learning criterion, uniform to all pairs of word nonsense syllables, was a single correct response. When a nonsense syllable was correctly anticipated, the corresponding card was eliminated from the series. Cards in a set were shuffled between presentations. When each of the nonsense syllables in a set had been correctly anticipated, to equalize the recency of stimulation between pairs correctly associated at an early presentation and those not anticipated until a later trial, all of the cards in that set were reassembled and again presented to *Ss* for a refresher trial. Following a rest period of 15 minutes *S* was required to relearn to the criterion of a single correct response the set of associates utilized in the original learning, except that upon this presentation of the set, *S* was afforded an opportunity to recall the nonsense syllable. The entire procedure provided data relative to number of presentations necessary for learning and relearning as well as number of associates correctly recalled upon first presentation in the relearning series.

RESULTS AND DISCUSSION

The learning, recall, and relearning scores were treated by an analyses of variance. Tests of homogeneity of variance revealed that in no case was it necessary to reject the hypothesis of homogeneity. Since it was probable that correlation would exist among the learning, recall, and relearning scores, analyses of covariance were also carried out.[2]

It will be recalled that the first hypothesis stated: "Associates to sexual words will be learned and retained less efficiently than associates to neutral or unpleasant words." The experimental findings do not bear out this prediction. Analyses of variance and covariance disclosed that the variance attributable to word lists and groups was short of significance. Since a consistent difference in favor of the nonsexual words was not revealed, the hypothesis was not confirmed.

The second hypothesis stated: "Anxious *Ss* will show a greater relative difficulty in learning and retention of associates to sexual words as compared with associates to neutral and unpleasant words than will normal *Ss*." Analyses of variance show the interaction, lists × groups, to be short of significance where learning scores are considered. However, in the case of recall and relearning scores, statistical significance was attained. The size of the interaction variance is reduced after treatment by analysis of covariance; however, it remains significant (at the .01 level) in the cases of relearning adjusted for learning and relearning adjusted for recall (at the .05 level). It very nearly attains the .05 level for relearning adjusted for both learning and recall. The large interaction variance which persists after adjustment by analysis of covariance suggests that the apparent interacting effect of the experimental variables is not only a function of differences of learning efficiency but actually operates during recall and relearning.

This interaction between word lists and groups is of the same nature in all the analyses. The form of the interaction is shown in Fig. 1 which is drawn from the relearning score means adjusted for both learning and recall. In all cases, the anxious group learned and retained the associates to the sexual words with greater efficiency than the associates to the nonsexual words, while the normal group learned and retained the associates to the sexual words less efficiently than the associates to the control words. Further analyses revealed that these differences resulted from differences in efficiency with most of the words of

[2] Tables giving data of the analyses of variance and covariance have been deposited with the American Documentation Institute. Order Document No. 4534 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 x 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

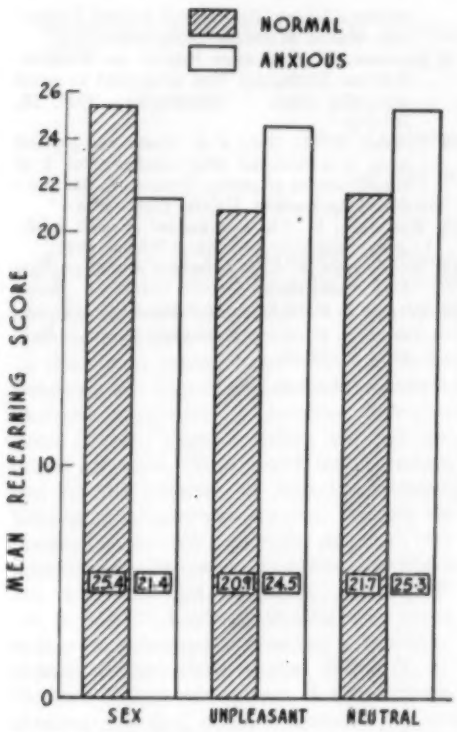

FIG. 1. MEAN RELEARNING SCORES AFTER ADJUSTMENT FOR BOTH LEARNING AND RECALL

a list rather than a large mean difference in the case of but one or two of the list. Since the interaction effect was consistently in the opposite direction to that predicted by the second hypothesis, the hypothesis was rejected.

The presence of the consistently large interaction variance and its similar nature in all analyses requires consideration. An *ad hoc* explanation may be suggested by the psychoanalytic concept of reaction formation. Reaction formation is anticipated only under special circumstances in the normal individual. It is conceived of as a loss of ability to repress which results from a change in personality structure that is marked by a neurotic tendency to react as if danger always threatens. The comparative efficiency of the anxious group with associates to sexual words with which the normal group was less efficient would seem to suggest the operation of reaction formation.

Finally, a question might be raised about the words in the unpleasant list. Because of the close relation of a number of words (i.e., rectum, excretion, bowels, urine) to various forms of perversion, this list could conceivably evoke the reaction primarily anticipated from sexual words. However, the mean scores on the unpleasant list are in every case closer to the mean scores on the neutral list than to the mean scores on the sexual list. Consequently, it seems likely that the observed interaction is ascribable to the meanings of the words rather than merely to differences in degree of unpleasantness.

SUMMARY

Freud's concept of repression was the source of two hypotheses: *Hypothesis I*. Associates to sexual words will be learned and retained less efficiently than associates to neutral or unpleasant words. *Hypothesis II*. Anxious *Ss* will show a greater relative difficulty in learning and retention of associates to sexual words as compared with associates to neutral and unpleasant words than will normal *Ss*.

Sixty normal *Ss* and 60 hospitalized anxious *Ss* participated in the experiment. The *Ss* were required to learn paired word-nonsense syllable associates to a criterion. The nonsense syllables were associated in various subgroups with the words of three lists—sexual, unpleasant, and neutral. Following a rest period the *Ss* were required to recall the nonsense syllable associated with the stimulus word and then relearn to the original criterion those pairs not correctly recalled.

The learning, recall, and relearning scores were treated by analyses of variance and covariance. These reveal no over-all significant differences between groups or among lists. Consequently, Hypothesis I was not confirmed. The interaction, lists × groups, was of a consistent pattern and in some instances at a significant level. The anxious group was more efficient in learning and retention of associates to sexual words than associates to neutral and unpleasant words, while the normal group was less efficient in the learning and retention of associates to sexual words than associates to neutral and unpleasant control words. Since the interaction was in an opposite direction to that anticipated, Hypothesis II was rejected.

REFERENCES

1. FREUD, S. Repression. In *Collected papers*. Vol. IV. London: Hogarth, 1949. Pp. 84–97.
2. FREUD, S. The unconscious. In *Collected papers*. Vol. IV. London: Hogarth, 1949. Pp. 98–134.
3. FREUD, S. *An outline of psychoanalysis*. New York: Norton, 1949.
4. FREUD, S. *A general introduction to psychoanalysis*. New York: Liveright, 1935.
5. FREUD, S. *The problem of anxiety*. New York: Norton, 1936.
6. RAPAPORT, D. *Emotions and memory*. Baltimore: Williams & Wilkins, 1942.
7. SEARS, R. R. Survey of objective studies of psychoanalytical concepts. *Soc. Sci. Res. Coun. Bull.*, 1943, No. 51.
8. SINGER, M. An investigation of the multiple choice projective test as a screening device to detect psychopathology. Unpublished doctor's dissertation. Univer. of Pennsylvania, 1948.
9. SPRINGER, N. N. A short form of the Wechsler-Bellevue Intelligence Test as applied to naval personnel. *Amer. J. Orthopsychiat.*, 1946, **16**, 341–349.
10. TERMAN, L. M., (Ed.) *et al. Mental and physical traits of a thousand gifted children*. Vol. I of *Genetic studies of genius*, Terman, L. M. (Ed.) Stanford: Stanford Univer. Press, 1925.
11. WECHSLER, D. *The measurement of adult intelligence*. Baltimore: Williams & Wilkins, 1944.
12. WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.
13. ZELLER, A. F. An experimental analogue of repression. I. Historical summary. *Psychol. Bull.*, 1950, **47**, 39–51.

Received March 20, 1954.

THINKING DEFICIT IN SCHIZOPHRENIA AND CHANGES WITH AMYTAL[1]

RITA SENF, PAUL E. HUSTON
College of Medicine, State University of Iowa
AND BERTRAM D. COHEN
Iowa City VA Hospital and State University of Iowa

A REVIEW of the literature on schizophrenic deficit indicates a need for more precise description of the deficit and the conditions under which it is elicited. In this report schizophrenic deficit in verbal reasoning will be analyzed in terms of several variables suggested by previous investigations. Briefly, these variables are: (*a*) sustained attention, (*b*) degree of task structure, and (*c*) the influence of personal reference. Kendig and Richmond reported that on the Stanford-Binet schizophrenic patients "fail conspicuously on tests of whatever type which call for prolonged attention and effort" (10, p. 73). On relatively unstructured tasks, such as the Rorschach and sorting tests, schizophrenic patients show marked deficit (1, 3). Bleuler's clinical descriptions of schizophrenic thinking (2) and, later, Cameron's (4) research, using incomplete sentences, emphasized the impairing influence of personal preoccupations.

This report concerns three tests: a short form of the Wechsler, the Army Alpha Subtest 3, and an incomplete sentences test. These tests were given under two conditions, without and with drugs (intravenous sodium amytal and amphetamine sulfate). These drugs frequently increase verbal responsiveness and bring the schizophrenic patient into better contact with his environment (7, 8, 11).

Both chronic and early schizophrenic groups were included to evaluate the effects of chronicity. Manic-depressive patients, depressive type, served as a psychotic control group and neurotic patients as a nonpsychotic control.

[1] From the Department of Psychiatry, College of Medicine, State University of Iowa. This investigation was supported in part by a research grant from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service. The authors wish to thank Dr. Alfred Jacobs, who assisted in scoring the tests, and Dr. L. P. Ristine, who arranged for the testing of chronic schizophrenic patients at the Mount Pleasant Mental Health Institute.

METHOD

Subjects

The patients included 24 chronic schizophrenic (CS), 22 early schizophrenic (ES), 21 manic-depressive, depressive type (Dp), and 21 neurotic (Nt). The mean ages in years for the four groups were: CS 39, ES 26, Dp 37, NT 30. The mean durations of illness were: CS 12 years, ES 1 year, Dp 9 months, Nt 6 years. For further descriptive data see (5).

The Tests

Wechsler short form. Form I (nondrug) and Form II (drug) of the Information, Comprehension, and Vocabulary subtests were given and scored by Wechsler's (15) criteria. Gerboth's (6) results suggest that with this sequence of administration there would be no significant practice effect. Scores on these subtests have been reported to deteriorate only slightly with age (15). However, to minimize the effect of age differences, each subtest weighted score was multiplied by five and converted to an IQ for the subject's (*S*'s) age. Inspection of mean subtest weighted scores and these IQ's showed the same features.

Army Alpha Subtest 3. This test of "practical judgment" contains 16 multiple-choice questions or stems, with three possible answers for each (16). On Form 5 (nondrug) and Form 9 (drug), the experimenter (*E*) noted the number finished in 90 seconds, which is the standard time limit; but the subject completed all items. Scores available are the number right in 90 seconds (90″ Alpha 3), total number right (Total Alpha 3), and total time for all items.

Incomplete Sentences. This test contained 11 stems ending with "because." The example was the same as that in the Alpha 3 instructions. The *E* asked for further explanation of unclear responses. The total score was the sum of scores on the separate items. Items 1–3 will be referred to as Impersonal; items 4–8 as Unstructured (for comparison with the same Alpha 3 items); and items 9–11 as Personal (to be compared with the Impersonal items).

Impersonal and Personal items. A sample of an Impersonal stem is: "Rocks sink in water because . . ." The personal stems were constructed by inserting a personal reference: "*My body* sinks in water because . . ." The corresponding Impersonal and Personal items were scored by the same criteria, and credits from zero to five assigned to responses.

Unstructured and Structured items. The stems for the Unstructured items were taken from Alpha 3. These Alpha 3 items are called Structured items. The names Unstructured and Structured refer to the differing response situations. For Unstructured items *S* had to

TABLE 1

MEAN SCORES FOR ALL GROUPS WITHOUT AND WITH DRUGS

Group	Wechsler Information IQ		Wechsler Vocabulary IQ		Wechsler Comprehension IQ		Alpha 3 No. Right in 90″		Alpha 3 Total No. Right		Incomplete Sentences Total		Structure Discrepancy		Personal Discrepancy	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Nondrug																
CS	109.0	16.2	106.0	13.8	96.2	18.4	5.7	2.6	13.5	2.0	29.1	8.0	8.7	6.2	1.6	3.1
ES	114.7	12.6	111.4	17.0	109.9	21.0	7.2	3.3	14.1	1.6	37.8	5.4	6.2	4.6	1.9	2.9
Dp	113.8	12.4	114.6	14.4	119.6	15.5	6.8	3.0	14.3	1.5	35.8	5.7	6.3	5.9	−0.3	1.3
Nt	113.6	13.9	115.4	11.8	121.7	15.5	9.0	2.8	14.6	1.2	38.8	6.0	5.7	3.6	0.4	1.9
Drug																
CS	106.0	11.6	111.7	13.7	101.7	18.1	5.5	3.1	12.1	1.8	34.3	7.5	3.4	5.6	0.2	3.3
ES	105.7	14.2	114.8	14.2	115.4	17.2	7.0	3.5	13.0	1.6	39.3	4.0	3.1	5.0	0.6	2.2
Dp	111.7	13.8	120.9	14.3	116.3	16.2	8.7	3.6	13.7	1.8	41.6	4.8	3.3	4.6	0.7	2.0
Nt	107.1	13.3	121.6	13.1	116.9	14.0	9.0	3.1	13.5	1.6	40.2	5.1	4.5	4.2	−0.4	2.3

Note.—The *N*'s were: Wechsler, CS 23, ES 22, Dp 20, Nt 20; Alpha 3, CS 23, ES 21, Dp 21, Nt 21; Incomplete Sentences Total and Personal Discrepancy, CS 20, ES 13, Dp 15, Nt 20; Structure Discrepancy, CS 19, ES 21, Dp 20, Nt 20.

supply the answer, while for Structured items his task was to select the correct one of the three choices provided. The Structured items were the stems from numbers 2, 9, 11, 12, and 14 of Form 5, and numbers 2, 3, 6, 11, and 12 of Form 9. These questions were reworded in declarative form ending in "because." Unstructured items received five credits if the response approximated the correct answer on Alpha 3; otherwise no credit was given. Structured items were also given credit of five or zero.

Procedure. The order of administration of the tests was, Wechsler, Incomplete Sentences, and Army Alpha Subtest 3, each given on different days, along with other tests of a larger battery. After an average interval of nine days, *S*s received sodium amytal and amphetamine sulfate intravenously and were retested. The average injection was 0.3 gm. of sodium amytal and 13 mg. of amphetamine sulfate. The drugs, mixed in the same syringe, were given until slight nystagmus appeared on lateral gaze. Testing began after any unsteadiness or lightheadedness subsided, usually within 15 minutes. Each *S*'s cooperation on each test was rated on a scale (9) from A to E. Only patients with C or better cooperation on both the nondrug and drug tests were included in the statistical analysis. Correlations of cooperation with the various test scores were low. Cooperation did not appear to differentiate the groups.

RESULTS

Wechsler short form. The mean Verbal IQ's were: nondrug, CS 104, ES 112, Dp 116, Nt 117; drug, CS 107, ES 112, Dp 117, Nt 115.[2] On the nondrug test the CS group was significantly lower than the other three groups (CS vs. ES, $p < .10$; CS vs. Dp and CS vs. Nt, $p < .01$). No other pairs of groups differed significantly. Differences between nondrug and drug tests were not significant in any of the groups. Since the four groups have approximately the same mean (12 years) and range of educational level, it seems reasonable to assume that the lower Verbal IQ's of the schizophrenic groups represent the effects of the psychosis, and that their prepsychotic levels of intelligence were equivalent to those of the nonschizophrenic groups.

On neither the Information nor the Vocabulary subtest (Table 1) was there any significant difference among the groups. With drugs all groups changed significantly, becoming poorer on Information ($p < .001$) and better on Vocabulary ($p < .001$). The magnitude of the change with drugs did not differ significantly among the groups on either subtest.

On the Comprehension subtest (Table 1)

[1] A table of means and *SD*'s for the Wechsler Short Form Verbal IQ, Army Alpha Subtest 3 total time, and the Structured, Unstructured, Impersonal, and Personal items has been deposited with the American Documentation Institute. Order Document No. 4441 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for 35 mm. microfilm or $1.25 for 6 × 8 in. photocopies. Make check payable to Chief, Photoduplication Service, Library of Congress.

without drugs the CS group had a significantly lower mean IQ than the other three groups (CS vs. ES, $p < .05$; CS vs. Dp and CS vs. Nt, $p < .001$). The ES group was also below the nonschizophrenic groups (ES vs. Dp, $p < .10$; ES vs. Nt, $p < .05$). The two nonschizophrenic groups did not differ significantly. With drugs the schizophrenic groups improved, while the nonschizophrenic groups became poorer. A t test showed that the schizophrenic groups did not differ from each other in amount of change with drugs; neither did the nonschizophrenic groups. A t test of schizophrenic vs. nonschizophrenic change with drugs showed that in these two larger groups, the change for schizophrenic patients was significantly different from that for nonschizophrenic patients ($p < .01$).

Army Alpha Subtest 3. For 90″ Alpha 3 (Table 1) on the nondrug test the Nt group was significantly better than the other three groups (Nt vs. CS, $p < .01$; Nt vs. ES, $p < .10$; Nt vs. Dp, $p < .05$). The other pairs of groups did not differ significantly. The drugs affected the groups differently ($p < .05$); there was essentially no change in any group except the Dp, which improved significantly ($p < .01$), coming up to the level of the Nt group. On the drug test, then, the schizophrenic groups were differentiated from the nonschizophrenic (CS vs. Dp, $p < .01$; ES vs. Dp, $p < .10$).

For total right on Alpha 3 (Table 1), on the nondrug test the groups did not differ significantly. With drugs all groups were poorer ($p < .001$), the largest decrement being in the CS group and the smallest in the Dp group.

On total time scores, with drugs the Dp group speeded significantly ($p < .05$), while the other groups did not change.

Incomplete Sentences. Without drugs, on Incomplete Sentences total score (Table 1) the CS group was significantly below the other groups (CS vs. ES and CS vs. Nt, $p < .001$; CS vs. Dp, $p < .01$). Among the other three groups no significant differences were found. With drugs, for all groups considered together, there was a significant improvement ($p < .001$). The amount of improvement was not significantly different among the groups. With drugs the relations among pairs of groups were the same as on the nondrug test.

Unstructured and Structured items. Table 1 presents results for the "Structure Discrepancy," the difference between Structured and Unstructured item scores. On both the nondrug and drug tests, all groups were significantly better on Structured than on Unstructured items ($p < .001$). The Structure Discrepancy did not differ significantly among the groups on either test. Nevertheless, by inspection the nondrug Structure Discrepancy was greatest in the CS group and least in the Nt group.

Drugs significantly decreased the magnitude of the Structure Discrepancy ($p < .001$), but not differentially for the groups. This decrease came about by a significant improvement on Unstructured items ($p < .01$), and a significant drop on Structured items ($p < .05$). The Dp group was an exception in that it did not drop with drugs on the Structured items. By inspection the reduction in Structure Discrepancy with drugs was greatest in the CS group.

Impersonal and Personal items. The results for the "Personal Discrepancy," the difference between Impersonal and Personal item scores, are given in Table 1. Without drugs, in both schizophrenic groups the Personal Discrepancy was significantly greater than zero ($p < .05$); that is, they were poorer on Personal than on Impersonal items. The two nonschizophrenic groups had no such Personal Discrepancy. With drugs there was no longer any significant difference among the groups in Personal Discrepancy; it can be considered zero in all. The removal of the Personal Discrepancy in the schizophrenic groups was due mainly to their improvement on Personal items.

DISCUSSION

Nondrug performance. The failure of Information and Vocabulary to differentiate among diagnostic groups is in agreement with previous findings (12, 13, 15). In both these tests, the attention required is minimal; the questions are quite simple and short. Correct answers are available through relatively simple associative processes rather than by complex reasoning.

The 90″ Alpha 3 score differentiated between psychosis and neurosis, but not among the psychotic groups. Shakow attributed the

lower performance of schizophrenic patients on this test to "the combination of conceptual thinking and social analysis required . . ." (14, p. 45). This test is self-administered, and the 90″ score depends on speed of response. On Total Alpha 3, with speed ignored, the schizophrenic groups were not impaired; the "conceptual thinking and social analysis" required by these items does not appear deficient. The schizophrenic deficit found on the 90″ score may indicate faulty maintenance of sustained attention. Shakow suggested that a defect in sustaining attention was not crucial on the 90″ Alpha 3, since his patients were unimpaired on the arithmetical reasoning and number completion Alpha subtests, which also involve speed and self-administration. What may be important is the interaction of the attention requirements with the social content of the Alpha 3 items.

In contrast to the multiple-choice structure of Alpha 3, the Incomplete Sentences test requires *S* to formulate his own answer. This relatively unstructured response situation apparently creates special difficulty for chronic schizophrenic patients. This interpretation is tentatively supported by the Structure Discrepancy results, and agrees with Cameron's finding that incomplete causal sentences are particularly difficult for severely disorganized schizophrenic patients (4).

The differentiation of the schizophrenic from the nonschizophrenic groups by the Comprehension test is consistent with other reports of the sensitivity of this test to schizophrenic impairment (12, 13, 15). This is usually attributed to a deficit in social or practical judgment. In addition, some of the questions are long, requiring more attention than Information and Vocabulary. The content of the questions is similar to that of Alpha 3 and Incomplete Sentences, but the response situation is less structured. Another requirement deserves emphasis. In many of the questions, the content tends to be personally involving; for example, "Why should we keep away from bad company?" "What should you do if . . ." Personal preoccupations aroused by such questions may interfere with reasoning in schizophrenia (4, 12). This test, then, sets many requirements which are difficult for schizophrenic patients to satisfy, and especially for chronic schizophrenic patients. The Personal Discrepancy results support the interpretation that schizophrenic patients are less capable of maintaining an objective orientation when the test questions contain personally involving stimuli.

On object sorting these schizophrenic groups were not adversely influenced by personally involving stimuli (5). Schizophrenic deficit may depend not only on the kind of personal stimuli used but also on the type of task into which they are introduced. In the object sorting test the emphasis was on placing objects into conceptual groups while in the present tests the emphasis was on appropriateness of verbal content.

Effect of drugs. On 90″ Alpha 3, the only appreciable change was in the Dp patients, who consistently speed under these drugs (Alpha 3 total time; 5, 9). Since on Total Alpha 3 the Dp patients showed no deficit, their poor nondrug 90″ score seems due to their slowness. The drugs may have a specific effect on psychomotor retardation, a prominent symptom of depression. The failure of the schizophrenic groups to improve under drugs either on the 90″ score or in total time suggests that their slowness differs from that in depression, possibly signifying fluctuations in attention rather than a consistent slow tempo (9).

The general decline in score on Information, Total Alpha 3, and the Structured items suggests that the drugs make it more difficult to be highly precise. On these tests the appropriate answer is highly specific; there are few or no approximations scorable as correct.

In contrast, Vocabulary, Incomplete Sentences, and the Unstructured items require little specificity of response. There is a wide variety of acceptable verbalizations, and increased verbal responsiveness is likely to improve the score. The drugs, by increasing *S*'s verbal productiveness and decreasing his concern for accuracy, permit him to verbalize creditable approximations which, without drugs, might be withheld.

The abolition of the Personal Discrepancy in the schizophrenic groups was brought about by their particular improvement on Personal sentences. It appears that under drugs schizophrenic patients are not so disturbed by words which have an ego reference. They can respond more objectively and produce answers

which more nearly coincide with the criteria of appropriateness set by *E*.

The improvement on Comprehension for the schizophrenic groups can be related to the decreased influence of personal preoccupations, as indicated by the Personal Discrepancy; and perhaps to an improved ability to cope with relatively unstructured response situations, as suggested by the Structure Discrepancy.

The lowering of Comprehension scores for the nonschizophrenic groups cannot be easily explained. It is possible that the high nondrug scores of the nonschizophrenic patients reflect their anxious concern to give socially approvable answers. With drugs they appeared more relaxed and less rigidly conscientious, permitting them to give less acceptable responses.

SUMMARY AND CONCLUSIONS

To study schizophrenic thinking, three tests (Wechsler short form, Army Alpha Subtest 3, and Incomplete Sentences) were given to four groups of patients: chronic schizophrenic, early schizophrenic, depressive, and neurotic. These tests were given without and with intravenous sodium amytal and benzedrine sulfate.

The nondrug results suggest that schizophrenic deficit becomes more apparent with greater requirements of sustained attention, with less structured response situations, and with tasks involving personal reference.

Two general effects of the drugs are suggested: (*a*) a facilitating effect when increased verbal responsiveness is beneficial to the score, and (*b*) an impairing effect when a highly precise response is required. In schizophrenic patients the drugs appear to improve reasoning by decreasing the effect of personal reference, which makes their responses more appropriate, and perhaps by improving their ability to cope with unstructured reasoning tasks.

REFERENCES

1. BECK, S. J. Personality structure in schizophrenia: a Rorschach investigation in 81 patients and 64 controls. *Nerv. ment. Dis. Monogr.*, 1938, No. 63.
2. BLEULER, E. *Dementia praecox or the group of schizophrenias.* (Trans. by J. Zinkin.) New York: International Universities Press, 1950.
3. BOLLES, MARJORIE, & GOLDSTEIN, K. A study of the impairment of "abstract behavior" in schizophrenic patients. *Psychiat. Quart.*, 1938, **12**, 42–65.
4. CAMERON, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
5. COHEN, B. D., SENF, RITA, & HUSTON, P. E. Effect of amobarbital (amytal) and affect on conceptual thinking in schizophrenia, depression, and neurosis. *A.M.A. Arch. Neurol. Psychiat.*, 1954, **71**, 171–180.
6. GERBOTH, RENATE. A study of the two forms of the Wechsler-Bellevue Intelligence Scale. *J. consult. Psychol.*, 1950, **14**, 365–370.
7. GOTTLIEB, J. S., & COBURN, F. E. Psychopharmacologic study of schizophrenia and depressions. Intravenous administration of sodium amytal and amphetamine sulfate separately and in various combinations. *A.M.A. Arch. Neurol. Psychiat.*, 1944, **51**, 260–263.
8. GOTTLIEB, J. S., BOBBITT, F. S., & FREIDINGER, A. W. Psychopharmacologic study of schizophrenia and depressions. VI. Differences in response to sodium amytal and benzedrine sulfate. *Psychosom. Med.*, 1952, **14**, 104–114.
9. HUSTON, P. E., & SENF, RITA. Psychopathology of schizophrenia and depression. I. Effect of amytal and amphetamine sulfate on level and maintenance of attention. *Amer. J. Psychiat.*, 1952, **109**, 131–138.
10. KENDIG, ISABELLE, & RICHMOND, WINIFRED V. *Psychological studies in dementia praecox.* Ann Arbor, Mich.: Edwards, 1940.
11. LINDEMANN, E., & MALAMUD, W. Experimental analysis of the psychopathological effects of intoxicating drugs. *Amer. J. Psychiat.*, 1934, **90**, 853–881.
12. MAGARET, ANN. Parallels in the behavior of schizophrenics, paretics, and pre-senile non-psychotics. *J. abnorm. soc. Psychol.*, 1942, **37**, 511–528.
13. RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing.* Vol. 1. Chicago: Year Book Publishers, 1945.
14. SHAKOW, D. The nature of deterioration in schizophrenic conditions. *Nerv. ment. Dis. Monogr.*, 1946, No. 70.
15. WECHSLER, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.
16. YOAKUM, C. S., & YERKES, R. M. *Army mental tests.* New York: Henry Holt, 1920.

Received March 23, 1954

PHONETIC SYMBOLISM IN NATURAL LANGUAGES[1]

ROGER W. BROWN, ABRAHAM H. BLACK, AND ARNOLD E. HOROWITZ

Harvard University

IN OUR experiments native speakers of English guessed the meanings of words from unfamiliar languages. The subjects (*Ss*) agreed very well on the meanings likely to be attached to these strange phonetic sequences. This agreement suggests a culturally acquired phonetic symbolism derived from experience with a common native language. However, the semantic judgments of our *Ss* were also correct more often than would be expected by chance, and this suggests that there may be some features of phonetic symbolism which have a universal validity. It is possible that speech originated in symbols imitative or somehow suggestive of their meanings and that traces of these "appropriate" linkages survive in all languages today.

NONLINGUISTIC PHONETIC SYMBOLISM

In 1929 Sapir (8) conducted experiments using pairs of "nonlinguistic words" built on the CVC model. Within a pair the two words were matched for the initial and final consonants while the vowels were varied in the range from "a" to "i" (e.g., "Mal"–"Mil"). Sapir asked his *Ss* to judge which member of a pair symbolized the larger reference. He found a strong tendency to consider the reference larger for the word that included a vowel nearer the "a" end of the scale.

In 1933 Newman (5) further analyzed Sapir's data and discovered systematic relationships between the size judgments and such factors as the position of the tongue in forming the vowels, the frequencies of the principal acoustic formants, and the size of the oral cavity created in pronunciation. Newman also discovered the magnitude implications of English consonants and investigated the symbolic values of the vowels and consonants for the dimension "bright to dark."

Bentley and Varon (1) used nonlinguistic words like those of Sapir and Newman together with some English words and standard nonsense syllables. Their *Ss* were simply asked to come up with a free associate for each item on a mixed list of such materials. The authors report no tendency to match reference magnitudes with nonlinguistic words in the manner that we should anticipate from the work of Sapir and Newman. From experimental variations Bentley and Varon learned that the reference value had to be forced out of the words by the suggestion of a scale of magnitude.

Bentley and Varon questioned the relevance of all of this work to the study of language. Both volume and brightness were known to be among the attributes of pure tones. Sapir and Newman might be said to have demonstrated that these same dimensions could be applied to complex vocal tones. Where in all this was there any justification for speaking of the "symbolic" value of "nonlinguistic words"? It had not been demonstrated that phonetic symbolism actually occurred in natural languages.

PHONETIC SYMBOLISM IN NATURAL LANGUAGES

In 1934 Tsuru wrote an unpublished paper called "Sound and Meaning" (10)[2] in which he described an experiment in phonetic symbolism. Tsuru compiled a list of 36 pairs of Japanese antonyms. His *Ss* were 57 Harvard and Radcliffe undergraduates who spoke English as their native language and disclaimed any knowledge of Japanese. Seated in a group, *Ss* first saw a pair of Japanese words printed in Romish characters. The English equivalent pair was then written on the board, the order of the words within the pair being randomly related to the order within the Japanese pair. From the back of the room a native Japanese pronounced the words twice, reversing the order the second time. The *Ss* then attempted to match English-Japanese synonyms with one minute allowed

[1] The authors are indebted to the Laboratory of Social Relations, Harvard University, for supporting this project.

[2] The paper was written for a psychology course with Professor Gordon W. Allport, who has kindly made it available to us.

for each pair. Table 1 indicates that they were able to average 69% correct guesses.

Tsuru had demonstrated that someone who knows both English and Japanese can find Japanese words whose meanings will be correctly guessed in a single alternative situation by *Ss* who know English but not Japanese. Although English and Japanese are believed to be historically unrelated, it would surely be possible to find some synonyms which sound alike. Tsuru's findings, then, could be consistent with the view that symbols are arbitrarily assigned in natural languages. If his *Ss* had learned in childhood to connect certain sounds in English with their arbitrary meanings they would be able to translate correctly Japanese words if these words were selected for their resemblance to the English. Tsuru's list does not appear to have been selected in this fashion, but the fact that he made his selection with full knowledge of both languages leaves this possibility uncontrolled. The work of Müller (4) with words from Swahili and Bantu suffers from the same difficulty.

Gordon Allport made the necessary improvement in an unpublished experiment.[3] He had a native speaker of Hungarian translate Tsuru's list into that language. This Finno-Ugric tongue is historically unrelated to either English or Japanese. Because the translator simply translated the list he was given, there was no opportunity to select words in the foreign language for their resemblance to English equivalents. The results appear in Table 1. Subjects performed somewhat less well than Tsuru's original group but were still above the chance level of 50 per cent.

In 1953 Susannah Rich (7), an undergraduate at Radcliffe College, wrote a senior honors thesis which bears in part on our problem. She used 25 pairs of words and had them translated into Japanese and Polish. Native speakers of the two languages recorded their pronunciations of the words. The *Ss* were 44 Radcliffe girls. The results appear in Table 1 and again are above the level to be expected by chance.

Neither Tsuru nor Allport conducted a test for the significance of the departure from chance represented by their results. Rich reported the mean percentages correct for both Polish and Japanese to be significantly better than chance ($p \leqq .005$). In general the results in Table 1 encourage the belief that there are significant traces of phonetic symbolism in natural languages.

TABLE 1

PERCENTAGES OF CORRECT TRANSLATIONS IN THREE STUDIES

Author	Languages	*N*	% Correct
Tsuru	Japanese	57	69
Allport	Hungarian	68	56.6
Rich	Japanese	44	57.2*
Rich	Polish	44	64.8*

* Reported to be significantly larger than chance ($p < .005$).

THE EXPERIMENT

English antonyms were selected from the Thorndike-Lorge word list with two considerations in mind. (*a*) The words should name sense experiences (e.g., warm–cool, heavy–light). (*b*) The members of a pair should both fall in the frequency range of 100 or over per million. The final list of 21 pairs consisted of those on which the three authors could agree. We were all completely ignorant of the languages into which the list was to be translated—Chinese, Czech, and Hindi.

The English lists were presented for translation to native speakers or scholars of the particular languages.[4] None of the translators was aware of the purpose of the study. They were asked to render the list into the familiar foreign equivalents—preserving the opposition within a pair. They then recorded their pronunciation of the pairs with five seconds between words and ten seconds between pairs.

The *Ss* were 86 Harvard and Radcliffe students in an undergraduate laboratory in social psychology. They participated in sections of about 20 students each. Our *Ss* were told that they were to take part in a study of sound symbolism in which meanings of foreign words were to be guessed from the sounds. Ranged down one side of a page were pairs of English antonyms. On the other side appeared foreign language antonyms matched pair by pair with the English. Words within a pair were randomly arranged so that while *Ss* knew that a given English word translated into one or the other of two foreign words he could not tell from the arrangement which of the two it was. The *Ss* heard the words pronounced in the same order as they appeared on the sheet and were then allowed ten seconds in which to match the two English words with their foreign language equivalents. The languages were presented in five random orders. All *Ss* were asked to

[3] ALLPORT, G. W. Phonetic Symbolism in Hungarian words. Unpublished experiment, Harvard Univer. 1935.

[4] The Czech was translated by Rose Adamek McGee, the Chinese by Dr. Bernard Solomon of the Harvard-Yenching Institute, and the Hindi by Dr. Surendra K. Srivastava.

describe their previous experience with the three languages. One *S* was disqualified by his knowledge of Chinese.

Preliminary analysis convinced us that the men and women did not show any significant differences in performance and consequently we combined their data. The results are presented in Table 2 in a form comparable to that used by earlier authors and summarized in Table 1. The mean percentages all depart significantly from chance. This calculation is made with reference to the binomial distribution. From these statistics we cannot tell whether all pairs on the list are judged with somewhat better than chance success or whether the average result is produced by combining a few words that are nearly always translated correctly with many that are at the chance level or below.

The results have therefore been computed in a different fashion. The performance on each pair of words for each language has been separately determined. We can then ask concerning each item whether more *S*s answered it correctly than would be expected by chance. These findings are summarized in Tables 3 and 4 as experimental condition A. On a given pair when there is agreement on 64 per cent of the judgments this result is significant at the .01 level. The great majority of the words in all three languages meet this criterion. It is clear that *S*s shared a conception of the more appropriate meaning for most of the words presented. We are interested in knowing how often this conception was correct. There must be 64 per cent correct judgments to reach the .01 level of significance. There are at least nine such pairs in all languages. There are only half as many significant results in the wrong direction. In other words our *S*s' conceptions of the proper translation were right twice as often as they were wrong.

The foregoing analysis describes the performance on each pair of words but it does not tell us how the performances of individual *S*s were distributed. Table 5 shows that most *S*s correctly translated over half the words in a given list. There were 70 *S*s for Chinese, 61 for Czech, and 67 for Hindi who had scores of 11 or better out of 21.

Allport's data have been subjected to the same kinds of analyses. Of his 36 pairs of Hungarian words 16 showed positive results significant at the .01 level; 53 of the 68 *S*s correctly translated over half of the words.

TABLE 2

PERCENTAGES OF CORRECT TRANSLATIONS FOR THREE LANGUAGES IN EXPERIMENTAL CONDITION A

Language	No. of *S*s	% Correct	Significance of Departure from Chance
Chinese	86	58.9	.001
Czech	86	53.7	.001
Hindi	85	59.6	.001

Possible Causes of the Obtained Results

How are *S*s able to arrive at so many correct translations? We suspect that the ability derives from a suggestive power in the phonetic sequences, but there are two other possibilities to be considered.

Differential length of words within a pair. Zipf (12) has demonstrated for samples of English, Plautine Latin, Peipingese Chinese, and several American Indian tongues that word frequency is related to word length, with the more common words tending to be shorter. It is possible that English words are correlated with their foreign language counterparts in both frequency and length. If this were so, an *S* need only match the longer English word with the longer foreign language word in each pair to achieve better than chance results. As a first precaution against this possibility we equated our English words for frequency range in the Thorndike-Lorge list. This step does not equate members of all pairs in length. Examining the data we find no indications that performance is better when the words differ in length, and, furthermore, we find that performance with the Chinese language, where the translations are all monosyllables of almost exactly equivalent length, is superior to performance on Czech, where the words vary in length.

The expressive quality of the speaker's voice. Our translators did not know the purpose of the study but were aware of the meanings of the words they pronounced. How can we be sure that the correct translations of our *S*s were not made possible by some expressive quality in the recorded voices? May they not have introduced some sharpness into their enunciation of "sharp," some haste into their versions of "fast"? Since these are sensory words it would be possible to adjust the qual-

TABLE 3

PERCENTAGE OF CORRECT TRANSLATIONS FOR EACH PAIR IN THREE LANGUAGES FOR EXPERIMENTAL CONDITIONS A AND B

		% Correct			% Correct			% Correct	
English	Chinese	A	B	Czech	A	B	Hindi	A	B
1. beautiful	meĭ	88*	70	krása	57	31	khubsurat	64*	50
ugly	ch'oŭ			ošklivost			badsurat		
2. blunt	tuň	78*	70	tupy	81*	83*	gothil	68*	83*
sharp	k'uăi			špičatý			tez		
3. bright	liang	67*	90*	svetly	64*	77	chamakdar	51	90*
dark	aň			tmavý			dhundhala		
4. coarse	ts'u	65*	70	hrubý	21†	44	mota	48	31
fine	hsi			drobný			achha		
5. down	hsiă	10†	31	dolů	56	50	niche	75*	83*
up	shăng			nahoru			upar		
6. drunk	tsui	66*	50	opilý	21†	70	nashe men	80*	77
sober	hsĭng			střízlivý			sanjida		
7. dry	kān	72*	70	suchý	44	50	sukha	42	44
wet	shīh			mokrý			bhiga		
8. fast	k'uai	83*	83*	rychlý	87*	83*	tez	27†	57
slow	màn			pomalý			sust		
9. fat	feĭ	31†	57	tlustý	69*	77	mota	66*	57
thin	shoŭ			tenký			patala		
10. gold	chīn	57	57	zlato	19†	57	sona	42	64
iron	t'iĕh			železo			loha		
11. bad	huai	34†	64	zlý	62	57	kharab	64*	31
good	haŏ			hodný			achha		
12. happy	huān	38	50	radostný	57	64	khush	17†	38
sad	peĭ			smutný			ranjida		
13. hard	kăng	97*	83*	tvrdý	76*	96*	sakht	61	64
soft	joú			měkký			narm		
14. light	ch'ing	93*	90*	lehký	66*	77	halka	36†	57
heavy	chuňg			těžký			wazani		
15. long	ch'ăng	55	44	dlouhý	80*	70	lamba	93*	70
short	tuān			kratký			chhota		
16. many	tō	73*	57	mnoho	55	25	bahut	88*	90*
one	yĭ			jeden			ek		
17. strong	ch'iáng	37	64	silný	28†	64	mazbut	34†	31
weak	jò			slabý			kamzor		
18. sweet	t'ién	58	51	sladký	24†	25	mitha	88*	70
sour	suān			kyselý			khatta		
19. thunder	leĭ	23†	31	hrom	92*	96*	garaj	62	77
lightning	tieň			blesk			chamak		
20. warm	nūan	73*	50	teplý	69*	77	garam	66*	77
cool	liáng			chladný			thanda		
21. wide	k'uān	37	90*	široký	43	57	chaura	76*	51
narrow	chaĭ			úzký			tang		

* Choices are correct with $p \leq .01$.
† Choices are incorrect with $p \leq .01$.

ity of speech so as directly to produce the sense quality in question.

To check this possibility we asked a control group of *Ss* (experimental condition B) to match words without the recorded pronunciation, using only the printed versions. These suggest a roughly correct pronunciation to most native speakers of English. The expressive quality of the voice is eliminated since *Ss* do not know the meanings of the words when they pronounce to themselves. The results for condition B, which appear in Tables 3 and 4, show that over-all performance in this control group was slightly more successful than in the experimental group. Correlating performances on individual items for the two groups we obtain an r of .59 for Chinese, .61 for Czech, and .55 for Hindi. All of these are significant ($p \leq .02$), and it seems to be clear that the words we have used have

TABLE 4

SUMMARY DATA OF TRANSLATIONS FOR 21 PAIRS OF WORDS IN THREE LANGUAGES IN EXPERIMENTAL CONDITIONS A AND B

Language and Experimental Condition	No. of *Ss*	Mean % Correct	No. of Pairs Significantly Correct	No. of Pairs Significantly Incorrect
Chinese				
A	86	58.9	11	4
B	16	61.9	5	0
Czech				
A	86	55.7	9	5
B	16	61.9	4	0
Hindi				
A	85	59.6	11	4
B	16	60.7	4	0

stable semantic implications which are not a consequence of expressive, nonphonetic aspects of pronunciation.

Interpretation of the Results

Since speech originated in prehistory, long before writing, there can be no record of the act or acts of origination. Traditional speculation on this subject has been trivialized with such titles as the "ding-dong theory" and the "bow-wow theory." The stale whimsy of this language has helped to make the subject distasteful. We shall avoid passing the standard theories in review, but will make one distinction among them. Some theories (especially Thorndike, 9) suppose that symbols are arbitrarily assigned to their references. It is then only necessary to assume the random emission of vocalization and some such learning principles as the familiar "contiguity" and "reinforcement" to explain the perpetuation and social diffusion of particular vocalizations. Other theories assume that the earliest speech, like the earliest writing (2), was "representational." There are three subvarieties of the representational position. The first two of these assume some kind of imitation. The onomatopoeia theory holds that vocal sound can suggest nonvocal sound. The gestural theory holds that motion and contour described by articulatory muscles can suggest motion and contour in the external world. Sir Richard Paget (6) has contended that human communication began with such gestural imitation of the external world. The third theory holds that vocalization falling on the auditory sense will naturally arouse meanings associated with other modalities. Hornbostel wrote of the "unity of the senses." Students of synesthesia have claimed (3) that perception is not an act involving only a particular receptor but is an affair of the whole body. Of course hearing does not involve all parts of the organism in equal degree but reducing hearing to audition alone means ignoring essential components. We doubt that this third variety of representational vocalization should be assimilated to synesthesia for the reason that studies of synesthetic response have typically yielded large individual differences. We prefer to use Werner's (11) term "physiognomic language" to name these universal, unlearned intersensory connections.

All of these speculations must be supplemented with a learning theory. They share, however, a belief that some sound-meaning connections begin with a little boost in habit strength that elevates them above the level of arbitrary connections. If languages all began by making use of these associations our data might be explained by the survival of some of these associations in modern languages.

The existence of onomatopoeia in modern, natural languages is widely accepted though it is considered to be far too infrequent to account for the origin of language. If we had selected our words for onomatopoeic possibilities no one would have been surprised at positive results. It would not be very interesting, for example, to demonstrate that subjects could distinguish the word for the cry of the rooster from the word for the cry of the cat in many languages. However, none of our words names an auditory experience and therefore onomatopoeia cannot account for our correct

TABLE 5

FREQUENCY DISTRIBUTIONS OF PERCENTAGES OF CORRECT TRANSLATIONS FOR THREE LANGUAGES IN EXPERIMENTAL CONDITION A

% Correct	No. of *Ss*		
	Chinese	Czech	Hindi
88			1
83	1	1	2
78	3	2	8
73	11	3	5
68	8	9	15
63	18	12	14
58	20	24	7
53	9	11	15
48	7	12	10
43	7	6	5
38	2	5	1
33		1	2

results. It is possible, however, that the origin of speech in gestural imitation and physiognomic language has left a residue of roughly translatable words in all languages and that we have tapped this residue in Chinese, Czech, and Hindi, while Allport tapped it in Hungarian, and Rich in Japanese and Polish.

The uniformity of performance among our *S*s might alternatively be explained by the fact that they are all native speakers of English who naturally share a cultural conception of the sounds likely to express a given meaning. The fact that our *S*s showed a significant amount of agreement in both incorrect as well as correct choices tends to substantiate this view. If the connections between linguistic sounds and meanings are arbitrary, however, the English speaker's cultural conception of phonetic symbolism should lead him to make correct translations of languages historically related to his native tongue but not to make correct translations of totally unrelated languages. Of the languages studied three are Indo-European (Czech, Hindi, and Polish) while three are not (Chinese, Japanese, and Hungarian). The Indo-European tongues should show a closer resemblance to English than the languages outside the family. However, there is no clear tendency favoring the Indo-European tongues. We are inclined to believe that most of our *S*s did not usually translate an English word with the foreign word bearing the closer resemblance, but rather responded in terms of principles of phonetic symbolism imperfectly realized in all languages. The words "light" and "heavy," for instance, translate into Chinese "ch'ing" and "chuŋ." As we might predict from the work of Sapir and Newman, 93 per cent of our subjects were correct on that pair.

Even if we assume a phonetic symbolism which is not a cultural acquisition, it does not necessarily follow that the presence of such symbolism in natural languages must be explained by the origin of speech in representational vocalization. Speech might have begun with the arbitrary association of sounds and meanings and then moved toward phonetic symbolism. A given speech form might have been aided in the struggle for survival if it chanced to be representational. Instead of assuming that speech has fallen away from a golden age of phonetic symbolism we may come to believe that, in the evolution of languages, speech forms have been selected for symbolism and that we are moving toward a golden millennium of physiognomic speech. Work is under way which we hope will decide between these interpretations.

SUMMARY

Three separate investigations, using three lists of English words and six foreign languages, have shown superior to chance agreement and accuracy in the translation of unfamiliar tongues. The agreement can be explained as the result of a "cultural conception" of the symbolic values attached to various phonetic combinations. This hypothesis does not explain the accuracy of translation. The accuracy can be explained by the assumption of some universal phonetic symbolism in which speech may have originated or toward which speech may be evolving. For the present we prefer to interpret our results as indicative of a primitive phonetic symbolism deriving from the origin of speech in some kind of imitative or physiognomic linkage of sounds and meanings. We forsake conservatism on this occasion for the excellent reason that the thesis proposed is so alien to most thinking in psycholinguistics that it needs to be brought forward strongly so that we may see that its unpopularity has not been deserved.

REFERENCES

1. BENTLEY, M., & VARON, EDITH J. An accessory study of phonetic symbolism. *Amer. J. Psychol.*, 1933, **45**, 76–86.
2. GELB, I. J. *A study of writing*. Chicago: Univer. of Chicago Press, 1952.
3. KOUWER, B. J. *Colors and their character*. The Hague: Nijhoff, 1949.
4. MÜLLER, H. *Experimentelle Beiträge zur Analyse des Verhältnisses von Laut und Sinn*. Berlin: Müller & I. Kiepenheuer, 1935.
5. NEWMAN, S. S. Further experiments in phonetic symbolism. *Amer. J. Psychol.*, 1933, **45**, 53–75.
6. PAGET, R. *Human speech*. New York: Harcourt, Brace, 1930.
7. RICH, SUSANNAH. The perception of emotion. Unpublished honors thesis, Radcliffe Coll., 1953.
8. SAPIR, E. A study in phonetic symbolism. *J. exp. Psychol.*, 1929, **12**, 225–239.
9. THORNDIKE, E. L. The origin of language. *Science*, 1933, **77**, 173–175.
10. TSURU, S. Sound and meaning. Unpublished manuscript on file with Gordon W. Allport, Harvard Univer., 1934.
11. WERNER, H. *Comparative psychology of mental development*. Chicago: Follett, 1948.
12. ZIPF, G. K. *The psycho-biology of language*. Boston: Houghton Mifflin, 1935.

Received March 23, 1954.

CASE REPORT

SOMATIC THERAPY AS DISCUSSED BY PSYCHOTIC PATIENTS

LIBBY BLEK

VA Hospital, Canandaigua, New York

AND LESLIE NAVRAN

VA Hospital, American Lake, Washington

THE past two decades have seen increasing emphasis in psychiatry on the use of several methods of "shock" treatment. Insulin therapy was the first of these to be introduced, in 1936, into American psychiatry. A few years later electric convulsive shock treatment came into use. For the theme of this paper it is not necessary to detail these treatments; suffice it to say that the methods are widely used and with good results in many cases.

There are various interpretations of the results of these therapeutic shock procedures (5). Our paper does not present an addition to past theoretical formulations per se; rather, the focus is on the psychological reactions of the patients. We shall present here the verbatim, tape-recorded account of an unrehearsed meeting of a group of patients at a neuropsychiatric hostpital. They were brought together to participate in a panel discussion on mental illness as part of a public information project. In the course of this meeting they discussed their reactions to somatic treatment methods.

There have been several publications dealing with the subjective reactions of patients to insulin and electric shock therapy. These include retrospective accounts by patients when they have attained a degree of remission (1, 2, 3) and examples of reactions by various patients given by the physician (4). In 1948 Alper offered as a basic reason for the lack of public case material written by patients that "... therapists have not been sufficiently aware of the potential usefulness of such personal documents ..." Also, she suggested that "typically the patient is too ill or too anxious, both before and after treatment to cooperate with the therapist" (1).

Our paper demonstrates that some patients are accessible enough to describe their subjective impressions even though they are acutely ill. Such reports aid in understanding the psychological aspects of the treatment process and the dynamics of mental illness. It seems to us that there is another equally important function to be served by studying subjective experiences, namely, how best to approach the mentally ill person. Those who prescribe and give the treatment should find it important to know the patients' feelings about how the treatment is given. The personal aspect was highlighted for the authors by one patient who, in describing and comparing shock treatments in various hospitals, said, "I've had 50 [EST]. I suppose that doesn't make me an authority, but..... The last doctor I had at B—who gave me shock treatment was a young man, and he seemed to be, well—cold. I mean he didn't seem to take an interest in the patients at all. I mean, he figured it was just another treatment. And it was hard taking shock treatments with him. With your doctor here, he takes more or less of an interest in you ... He's friendly with you and he talks with you. He asks you how you feel. This man would just say 'Next,' and you would go in and sit down, he'd hook you up, and you'd get the shock treatment. Even the aides he had felt a little funny about it."

Our purpose, then, is to present a transcript of a panel discussion by four patients who gave us their permission to present their reactions for scientific purposes. Treating the mentally ill person is a joint venture of patient and physician and in the discussion which will follow the patients' statements, we will attempt to focus not only on the patients' reactions to past treatment but also on some of the implications of these reactions for hospital routines.

THE PATIENTS

Of the four patients whose subjective reactions are presented here as they shared them

with each other and the authors, two patients —W and C—have since left the hospital, their psychoses in remission. L and D remain in the hospital and are seriously disturbed people, so much so that they spend most of their time on the maximum security ward. At the time of our meeting, W was 36 years of age, L was 25, C was 32, and D was 24. Rather than continue with a short case history of each patient, we believe that the reader will be able to get a better picture of what each patient was like as an individual, and the psychological level at which he was functioning at the time of the interview, if we let each one introduce himself and describe his own illness as he did in the panel discussion:[1]

W. I was turned in here from the Navy. I was working—doing my job—but I think I was working too hard. I think that I was overworked, concentrating on my job. I had recently come out of Korea. In addition to that I had family trouble and the load got too much for me to carry. I—therefore, I'd taken sick—turned into the doctors at sick bay. By the time I turned in I was sick, desperately sick. I needed treatment of some sort. I ran into a blank wall and stopped. I couldn't go any further. The route I was taking at the time I turned in, I would have been committed or I would have been dead. What made me know I was sick—because under ordinary circumstances I have a certain amount of self-control—I don't mean I've got steel nerves or anything like that, but I don't usually fly up when I have things said to me because I have heard practically everything anyone can say.[2] On this particular occasion I was having family trouble. I went to talk to the man about the trouble I was having. He called me names and I flew off the handle. I mean I went all to pieces. In fact I threatened to fight him in his office. That was the signal to me to know that there was something wrong with me because I had been called that particular name before and I heard people called even worse names than that. On this particular occasion I blew up and threatened to fight him. I threatened to kill him in fact—just for calling me a name and that was the signal that something had gone wrong. I got to the point when a person couldn't say anything to me unless I flew off the handle and got ready to fight and do violence—well, then I knew I was sick.

L. I would like to say a little bit regards myself and my illness. The root of my illness originated with devoting too much time to things like theology. I just lost track of myself and didn't have anything to tell me when I went far enough or whether I'd went far enough. I knew for a fact I was slipping—I'd lost complete contact with nature, trees, or the outside. Being outside or anything like that didn't interest me at all. Everything was print to me—the only thing of any interest was print.

C. A man knows when he doesn't feel good. You can feel yourself backsliding. How about if one or two doctors had diagnosed your case? They've made a mistake—yet, say you're committed and you go to a hospital and to another hospital and you're getting worse and you keep protesting that this hurts you, it bothers you and you want something done and you're getting worse. And yet you can't get anything done. What would you do? All a man asks out of life is a chance. Before I go home I'd like to sit down and talk to the doctors and go over my complete case history so I'll know exactly.

D. I was put in the hospital as a result of a conflict at home with my father. My father felt that I should be in a mental institution and I resisted him—and got angry with him—and the display of emotion frightened him and he called the police and had me committed. Before I was committed I was very nervous. My behavior was erratic and it is true it caused concern to my parents. It is true that I wasn't up to par emotionally—perhaps it's a pipe dream, but I really wish that there could have been another way to get around it. I had the feeling that things were against me. It was accompanied not too much later by shakiness, loss of balance, and inability to stay in public places without feeling great fear. And with the onset of these symptoms I knew that I was mentally ill.

THE PANEL DISCUSSION

Doctor. Now then I'd like to ask you if you would like to talk about the kinds of treatments you've had here. And you can discuss them any way you want to—if you like them, dislike them, so forth and so on. Say what the treatment was and tell us how you feel about it or how it helped you. The lid's off.

W. That depends on how far from the future the treatments and I are separated. If I look back years later and see that I'm a different individual and that I've adjusted myself to living, then of course I will eh, speak, you know, on the treatments. But right now I wouldn't say. I wouldn't want to say because I haven't had a chance to test the treatments. See what I mean? Now two or three years from now when I'm away from it and I find that living is easier for me than now—of course I could say on the treatments.

Doctor. All right, that's a fair answer. Anybody else?

L. I think that shock is one of the outstanding treatments here at the hospital. I have had quite a bit of it myself and I know I have been aggravated on the wards to the point where I just—would ask the doctor if I could go on

[1] There has been a minimum of editing. In this introductory section, where statements by one patient were interspersed with personal data remarks by another patient, the remarks of each patient were united so as to give a clearer picture.

[2] This patient is a Negro.

the shock table in the morning in order to get away from it and get some rest. That sounds like a horrible thing to request to some of the patients, but it really helps me.

C. I don't feel I'm qualified to speak on the different treatments as to what they've done to my system because here again as I've said I'd like to talk to the doctors or to a particular doctor and find out just what has been done and what it has done to me supposedly and go over my case history. I do have this lapse of memory. You can't just shrug your shoulders and say it's just a blank spot in my life because after all it's an important part of my life and has a bearing on my future and I don't feel qualified to speak.

Doctor. I'm really not asking you to make a medical evaluation of the good of the treatment. I just want your own frank reaction. First of all, I'd like to know what kind of treatment you had and then how you felt about them while you were having them.

C. I can't say that I like shock treatment. I'd rather do something else than take a shock treatment. I realize though when the doctor does give a treatment you are doing all you can to help us here and anything we can do to cooperate we are going to do.

L. What hurts me most about my treatment is not taking the treatments itself—it's being told by people in charge "OK let's go" and things like that. I'd rather walk to the shock room myself than be escorted by two guys that have the attitude that says, "We'll give you shock right here if you don't go."

D. I agree with Mr. L about the psychological effect before you get a jolt. It's always made me weak in the knees and I know there's been a couple of times when they have had to sit on me six and seven to make me take it. Sometimes I thought that the fear before shock outweighed the value after taking it.

Doctor. We heard some expressions from D just now about the feeling you can have before you get the treatment. L was saying he wants to take treatments because he thinks they help him and C and D are saying that it isn't particularly pleasant. What about the actual treatment itself? What do you experience when you get an electric shock treatment? Do you feel the jolt at all—or what?

W. I don't have time to feel it. There's no pain to it. Black right out.

D. I think Mr. W is right. It's too quick. I don't know what the time element is.

W. Before you take it it's terrifying to think of having to go and take it.

D. I have known fellows who can just walk in and just strap on the electrodes and wham. In fact I know one or two who have said they didn't mind it and enjoyed it.

W. I've heard that too. I've heard it, but I've always doubted it.

C. If I ever have any more to take, well, they can take my share of them.

L. The thing I experienced—if you can think of something else besides what you are taking. If you devote your time to what they are going to do to you, it's going to do more harm to you than the treatment itself. I just—I think, "I wish I was out flying," and glance out the window in back—

D. (cuts in) Before you know, you are flying.

L. There is two ways of giving them. At least I think there is. When I first took them it seemed to me they were giving them to me gradually. It didn't knock me out instantly. I don't think they gave me as much electricity as they did later on. I used to—I remember myself hollering and some one trying to bear down on me and they used to have to strap me to the table and everything else in order to give me one. It lasted a few seconds before I was completely out. I've taken the other kind too, where you just say "OK, open your mouth, bear down on the plunger." Then I glance out through the window in back of me and wish I was flying, and I wake up drinking coffee.

Doctor. What about waking up? How do you feel?

L. Some mornings when I have taken the electric shock and came out of it, I found myself strapped feet and hands. My hands were turned loose and I had my coffee and toast and a few minutes later, why they let you out of bed and you're back into your everyday clothes—and occasionally I have difficulty finding out where I came from. I'm relaxed awful mentally but my past is completely abolished for at least an hour or two afterwards. I would have to ask—like when I was on Eight building, I would have to go to the attendant to find out which ward I

came from. There are four wards on the building and I would have to find out which one I came from.

D. I've noticed after my shock treatments the returning of my memory to very early years of life. The second shock treatment I took, uh, when I awakened from it I believed I was about eight or ten years old. I was twenty actually, but on awakening I felt that I was a boy again and that it was morning and time to go to school and the world seemed very pleasant—there were no difficulties and anxieties. That condition lasted for quite a time before a nurse came to the door and brought me back into my environment. It is the strangest reaction I ever had from anything.

Doctor. Anybody else?

W. Well, about a lapse from memory I would elaborate on it by saying that after taking shock, some things I forgot altogether and don't know anything about it. I forgot when I was married for one thing—which is important—and I forgot how many children my sister has—I'm saying that to emphasize that I forgot everything. A lot of things came back to me by reading and association. By reading and association with other people. For instance, I had a visitor only recently. He came out and talked to me and brought things back to my mind that happened three or four years ago that I had completely forgotten. I had forgotten completely all about it—trips, places we had gone together, so that still holds good with me. I still have certain things that I've forgotten—when it's ever mentioned to me it comes back just like that.

C. Well that's where you're fortunate, W. I've got an album home of a lot of pictures and places I've been, things I've done. I've gone through the album and I've seen pictures of myself taken alongside a sign of Fort Knox, Kentucky. I don't remember Fort Knox. That's where you are fortunate to remember back.

W. It comes back when it's called to my attention.

C. That's one of the reasons I want to get back some of those blank spots if I can. But just the important things. The other little things, that's all right, I can take that in my stride. But the important things that I need to remember, like the fact of making a living, getting along in the world, that's the thing I'll have to remember.

L. I want to put emphasis on what I said in regard to memory after the shock treatment. It's just that it's abolished for a short period of time after the treatment. Then it comes back to me—I don't have to be reminded of it.

W. If it's a long association in the past I have to be reminded of it. I still stand by that. If it's recent I can remember it OK and after it's called to my attention once, it come back for all times. Something that happened four, five years ago just like C said, it's completely gone. I don't know anything about it all, but if it's called to my attention I believe it will come back.

Doctor. I think that what we have been saying is that electric shock certainly does affect your memory for a longer or shorter length of time, and that there are individual variations as to how much it affects your memory. I think right in our own group we've been saying that it affects each of us differently—What other kinds of treatments have you had?

W. I have had insulin.

Doctor. Do you want to tell us something about that? What you felt while you were taking it, and so on.

W. I don't know exactly. Coma, seems—it frightens me. Other than it frightens me. I don't know how to express it.

D. I agree with Mr. W that insulin seems like a pretty tough go. The coma that insulin produces has been a scary thing to me anyhow.

W. One thing I think why insulin frightens me so is that I do things that I wouldn't do ordinarily. I fight the attendants, I say things that I will say but I don't say them in the way—I mean—use profane language for instance. In fact two or three hours after the treatment, in addition to being awfully tired, I'll say things that will sound irrational to me. I find myself sounding like some of the other patients that I don't want to sound like, even were they at their so-called best. I believe that if it wasn't for the fact that I was going into a coma not knowing what I was saying or doing, that I wouldn't be so frightened. But since I'm completely gone into a coma, why it makes me—it frightens me.

L. I had insulin when I first came here and

the thing that bothered me about it was the needle itself. My body was terribly run down. I was frail and my goodness when that nurse stuck that needle in me why—I don't think it would have hurt so bad if I had been a little huskier. But probably if I had been a little husky I probably woudn't have needed the needle. Those needles, my goodness! They hurt horribly.

C. A fellow does feel like a pin cushion. I'll go along with you on that.

Doctor. What happens when you get an insulin treatment? You get a needle put into you, L?

L. It seems to me it was in the hip. I don't remember now. I took it when I first came and I have been here almost thirty months. I do know the needle hurts. That's the outstanding thing about it, they hurt.

Doctor. I wonder if anyone else can tell anything about it? How it is done? What do they do to you?

C. They start off by giving you a shot in the arm. They build up the dosage and start in the hip and they give it in both hips. Like L said, you feel like the "target for tonite." But like I have said before, we realize these things are done for our good and anything we can do to help, naturally—.

Doctor. What happens after the needles?

C. You lay there for a while and they want you to go to sleep and I try to go asleep. That's what I generally do. I get awful thirsty, and wringing wet with sweat, and then you feel like you're going to crap out and you have had it, and then I ask to be strapped to the bed, and the next thing you know—later I'm brought to with a shot in the arm, or somebody is saying drink this, and your head is going round and round. That's about all I can say.

W. About this sleeping. Periodically the nurse and attendant come to take your pulse. You may try to go to sleep, 'cause I tried myself to rush myself to sleep and tried to get it over with, but as fast as you just doze off there's someone will walk in and feel your pulse. You just can't relax. Soon as I get the needle I like to go to sleep. I would like to sleep from the time I get the needle until it's all over, but these periods when they take your pulse keeps you awake. It does me anyway.

D. Sometimes it seems to me, doctor, the only thing insulin does is put weight on a guy.

W. I don't think—go ahead—I'm sorry, go ahead.

D. Have you gotten a lot of beneficial—?

W. I seem to think the same way. I can't tell the difference in my attitude and I would like to ask are you supposed to think differently after taking insulin? Are you supposed to react differently? I would like to know because the only benefit I actually see is being built up. I have taken the treatment and two or three hours later I'm still W. I'm back to myself the way I was before.

Doctor. Is there any difference when you take the shock treatment? Does that change your attitude?

W. Not necessarily. It does black out everything.

Doctor. Why don't we let it go this way? One of the things you think about while you're taking insulin is: What is it supposed to do for me? You wonder what the effect was supposed to have been except possibly other than that of building you up. All right. How about the others in the group? Do you have the same feeling or do you have a different feeling? What does insulin do—does it make you different, or what?

C. I imagine we all wonder what it does for our bodies and what the treatment is, but I imagine it's the doctor's place to—.

L. I didn't—pardon me, C.

C. Uh-huh.

L. I didn't think there was any connection between insulin and the mind. I thought the purpose of insulin was to build the body.

C. One thing I couldn't follow you. You said something about after the treatment changing your attitude? Well, in my attitude I can't see where it needed any changing. I had some definite ideas on the future, but they'd be the same regardless.

W. That's the point. Why is insulin given to you? Why? Why?

Doctor. Actually, that's a question I'm going to duck except to point out it's a question you have. I can give you an answer, something that goes: You got it because the doctor thinks it's going to help you. That doesn't tell you anything, though. Still, to do other than that would be to go into a technical discussion of the effect of the insulin on the blood sugar content and so on, and I don't know that that would help you. For our purposes here, I'd

like to restrict my participation to asking you the questions rather than answering. That's a nice way of saying I won't answer right now. The point is the accent here really is on your feelings about your experiences, rather than making this a lecture where you would be learning something of technical information. That's why I've been consistently ducking the questions you have been asking me.—We have mentioned electric shock and some discussion of insulin. How do you feel when you come out of an insulin treatment? C has told us how he feels and I wonder if some of the others who have had insulin could talk about that. How did you feel when you came to?

D. I recall one treatment—having a terrific nightmare before waking up, doctor. It was really a terrible nightmare. And I knew another guy that would try repeatedly to run away rather than get caught up in insulin treatment.

C. You get a bad headache sometime.

W. You feel tired. I feel like I'm not able even to get up. After I'm unstrapped I don't know how I get up. Physically, I don't feel able to walk again from here to there. After I start and take a shower, gradually I start feeling strength coming back—but all my strength is completely gone.

Doctor. Anybody else got something to say on that?

C. I have the same feeling. I'd sit up in bed, sometimes I'd lie back over again. Like I say you are pretty tired.

Doctor. About how long would it take before you'd feel like yourself again?

D. I think as soon as you get some sugar in your stomach you start feeling yourself. Soon as you have eaten a full meal or had your glucose or whatever it is they pour down you—it seems to me you come right back out of it then.

W. Sometimes till the next treatment is the way I feel. From the time I come out till I get the next treatment after all night's sleep and exercise, then I start feeling like myself again. It takes about a day, about a day.

C. Some days I feel better than others. Some days I don't feel good at all. Just like with W, I can go all day until the next treatment.

L. I was going to say the same as Mr. D over there. Usually after your meals, the rest of that day why you began to feel yourself again.

DISCUSSION

As the panel discussion progressed and we listened to these four men describe their treatment experiences in simple, direct fashion, we were impressed by what they were telling us. Here we might point out that what they didn't say seems as significant as what they did say. For example, no one had experienced at our hospital what another patient had previously experienced in a hospital handicapped by lack of funds and personnel. He said, "*There* was a beautiful place for shock treatment. They formed you in a line, about forty in a line, uh, forty men. And the only thing separating you from the shock treatment was a curtain. And you could hear the men getting treatment before you. You just went right along, one right after another and you got all the sound effects. Boy, that registered!" Today, most of us would agree that we cannot divorce the actual giving of the shock from the experiences of the patient prior to and after the actual somatic treatment. We have become aware of the importance of the therapeutic orientation of all who come into contact with the mentally ill person. This brings us to a serious consideration of our patients' discussion and how we can use what they have told us to the best advantage for all patients.

All four patients had undergone the same somatic treatments, but their reactions document their individual variations on several important common themes. Thus, for example, they have memory difficulty, but some forget recent happenings, some the events of previous years; for some the lapse of memory lasts a few hours, for others it lasts for days. Again, they are resentful, but some resent the way treatment was given, some the attitude and behavior of the personnel with whom they are brought into contact, and others resent the treatment itself. They are chagrined, but some are chagrined at hearing themselves speak irrationally, some at their inability to understand the purpose of the treatment. Most important of all is the common reaction of fear. This fear is not aroused by physical pain—on the contrary, it seems to be centered in their expectation that a force is to be applied which they do not know anything about, but which

will have disturbing effects of which they will be keenly aware. It is our contention here that the apprehension and helplessness that lie behind W's cry, "Why? Why?" can hardly be considered therapeutic. But more important, these feelings make it seem crucial that all who come into direct contact with patients give some thought to ways of minimizing the nontherapeutic emotional reactions associated with current somatic therapeutic procedures.

Several possibilities suggest themselves, none of which is in itself a final answer. As a beginning, it would seem important to indoctrinate all personnel with a therapeutic orientation which would eliminate verbal forms of coercion which threaten the use of somatic therapy as a punishment. Another goal would be the reduction of physical handling of patients to the minimum necessary for the carrying out of the treatment.

Perhaps of greatest importance would be the provision of opportunities for groups of patients to discuss with each other and the doctor their fears and questions about treatment which they will receive or are in the process of taking.

The rapport that could thus be established between personnel and patients would grow out of the feeling that others are interested and do care. The therapeutic gains of such communication could do much to transform the entire hospital milieu in which somatic treatment obtains into an integrated situation which may enable other patients like our Mr. W to "look back two or three years from now . . . and find that living is easier for me . . . ."

REFERENCES

1. ALPER, THELMA G. An electric shock patient tells his story. *J. abnorm. soc. Physchol.*, 1948, **43**, 201–210.
2. ANONYMOUS. Insulin and I. *Amer. J. Orthopsychiat.*, 1940, **10**, 810–814.
3. FRASIER, R., & SARGENT, W. The subjective experiences of a schizophrenic illness: personal records written at the end of the illness by some patients who were treated with insulin. *Charact. & Pers.*, 1940–41, **9**, 139–151.
4. MOSSE, E. P. Electroshock and personality structure. *J. nerv. ment. Dis.*, 1946, **104**, 296–302.
5. STAINBROCK, E. Shock therapy: psychologic theory and research. *Psychol. Bull.*, 1946, **43**, 21–54.

Received March 20, 1954.

CRITIQUE AND NOTES

EARLY LANGUAGE BEHAVIOR IN NEGRO CHILDREN AND THE TESTING OF INTELLIGENCE

BENJAMIN PASAMANICK AND HILDA KNOBLOCH

The Johns Hopkins University

MUCH has been written concerning Negro-white differences in growth, development, and intelligence. With continued investigation, environmental causes for these differences are being discovered so that it is no longer possible to state that they are unequivocally innate or racial (3).

The most controversial area, that of intelligence, has been studied intensively and the responsible sociocultural variables fairly well delineated. Among these has been described awareness of race, even in the preschool child, with consequent loss of rapport and lowering of score on psychometric examination. Poor performance on verbal items, attributable to cultural factors, has also been implicated (1).

In a longitudinal study of a low socioeconomic Negro group in New Haven, no significant Negro-white differences in either physical growth or behavioral development were found during the first eighteen months (7). It was hypothesized that growth and development were causally related and that poor nutrition rather than innate inferiority was therefore responsible for differences previously reported in both growth and development of Negro infants (6). A third examination of the group (40 children) at the mean age of 24.4 months supported this hypothesis (4). Growth and development for the group as a whole continued at white rates.

With continued use of the Gesell Developmental Examination on this third examination, interesting findings were noted in two behavior areas. Gross motor behavior was found to be markedly accelerated. Language behavior of the Negro infants while not retarded (mean DQ = 101) when compared to the white norms was significantly lower than the other fields of behavior. In addition, it was lower than the scores achieved in the first (DQ = 106) and second (DQ = 104) examinations, the difference between first and third being statistically significant ($p < .01$).

In a recent and still unpublished study done by one of the authors with Williams and Scott (8), a probable explanation for the discrepant scores in gross motor behavior was found. Gross motor development in a series of Negro clinic and private patients was found to be negatively correlated with socioeconomic status, the latter being negatively correlated with permissiveness. This probably accounts for the high quotients in the New Haven group which is largely of low socioeconomic status.

In an attempt to explain the comparatively low scores in the language area, the hypothesis was advanced that awareness of the examiner's different skin color caused sufficient inhibition to decrease verbal responsiveness by the children. This point had been raised by one of the parents who offered it as an explanation for her child's almost total lack of speech during examination.

This hypothesis was tested by somewhat crudely dividing the language behavior items on the Gesell Developmental Examination into three spheres, reported language behavior, comprehension of language, and verbal responsiveness. This was done to compare verbal responsive behavior, which might be impaired, with reported and comprehension behavior. Language maturity levels were assigned on the basis of recorded schedules in each of these spheres and quotients derived by the formula $DQ = \frac{CA}{MA}$ (Table 1). The mean quotients arrived at in this fashion were quite revealing.

Verbal and comprehension behavior spheres are significantly different ($p < .001$). Reported behavior is not significantly different from the other two. It is also apparent that verbal responsiveness varied more ($SD = 17.7$) than did comprehension or reported behavior; this might be due to variation in awareness of racial differences.

Although reported language behavior fell between the other two spheres and was essentially the same as total language behavior in mean score, it was thought advisable to test accuracy of reporting by mothers. This was done by comparing

TABLE 1

MEAN DEVELOPMENTAL QUOTIENTS OF VARIOUS ASPECTS OF LANGUAGE BEHAVIOR

Language Behavior	Mean DQ	*SD*
Total	101	10.8
Reported	101	12.6
Verbal responsiveness	94	17.7
Comprehension of speech	104	15.7

the reported and observed gross motor behavior in the same manner. The mean reported behavior quotient was 131 and the observed 132; these are of course not significantly different indicating that, on the whole, the mothers of our subjects were fairly reliable reporters of gross motor behavior. There is no reason to believe their reporting of language differed from the actual performance of the child.

DISCUSSION

The hypothesis that racial awareness plays a significant role during the examination process for the assessment of development seems to have some validity, even in the early preschool years. Comparative studies of ethnic group and racial psychologic differences by investigators from backgrounds and with physical characteristics differing from those of the group under investigation are inherently biased by this variable. One could speculate as to the source of these difficulties. How much of the observed reserve and inhibition is due to physical differences in the examiner and how much due to learning of attitudes from parents and siblings? In any event, it is a factor to be reckoned with in projected studies and in the evaluation of past studies.

Another implication from these findings is the difficulty of interpreting differences found when verbal response items are used in intelligence testing. Investigators have noted these sources of error (2), and it is of interest that our findings show that they may arise at a surprisingly early age.

Landreth and Johnson have shown that by three years of age Negro children have learned and accepted the majority stereotypes about Negroes (5). If our interpretation of the cause of lack of verbal responsiveness in the Negro children is correct, the awareness of racial differences apparently occurs much earlier than has been previously demonstrated.

SUMMARY

Forty Negro children examined by a white examiner were found to have lowered language scores on third examination at two years of age. This was apparently due to lack of verbal responsiveness, rather than poor comprehension of language. This apparent early awareness of racial differences and loss of rapport has serious implications in the field of ethnic group psychology, particularly in the use of verbal items on intelligence testing.

REFERENCES

1. ANASTASI, ANNE, & D'ANGELO, RITA Y. A comparison of Negro and white preschool children in language development and Goodenough Draw-a-Man IQ. *J. genet. Psychol.*, 1952, **81**, 147–165.
2. BROWN, F. An experimental and critical study of the intelligence of Negro and white kindergarten children. *J. genet. Psychol.*, 1944, **65**, 161–175.
3. KLINEBERG, O. Race differences: the present position of the problem. *Bull. World Fed. Ment. Hlth*, 1951, **3**, 3–11.
4. KNOBLOCH, H., & PASAMANICK, B. Further observations on the development of Negro children. *J. genet. Psychol.*, 1953, **83**, 137–157.
5. LANDRETH, C., & JOHNSON, B. C. Young children's responses to a picture and inset test designed to reveal reactions to persons of different skin color. *Child Develpm.*, 1953, **24**, 63–80.
6. MCGRAW, M. B. A comparative study of a group of Southern white and Negro infants. *Genet. Psychol. Monogr.*, 1931, **10**, 1–105.
7. PASAMANICK, B. A comparative study of the behavioral development of Negro infants. *J. genet. Psychol.*, 1946, **69**, 3–44.
8. WILLIAMS, J., SCOTT, R., & PASAMANICK, B. Unpublished data available at Howard Univer., Wash., D.C.

Received March 18, 1954

RESPONDENT REACTIONS TO AMBIGUOUS ITEMS IN AN ATTITUDE SCALE

ROBERT O. BLOOD, JR.

University of Michigan

SANFORD and Older's "Short Authoritarian-Equalitarian Scale" (a revised form of the California F Scale) has been used by various social scientists because of its brevity and its relationship to the research on authoritarian personality structure (3; 1, pp. 248–250). Its administration by the writer to 80 middle-class, well-educated subjects living near Chapel Hill, North Carolina disclosed ambiguities which limit the scale's effectiveness.

For example, respondents are asked to choose one of six responses ranging from "agree very much" to "disagree very much" for Item E: "Husbands should help their wives with the dishes and care for the children." So many of our respondents objected to the double-barrelled nature of this item that we asked 24 of them to respond (*a*) to the above phraseology, (*b*) to the reference about dishes only, and (*c*) to the reference about children only.

Analysis of the results shows that most of the group responded differentially. Only six of the 24 responded uniformly to the three alternative forms of the item. For the remaining 18 interviewees, the responses to one or to both separate "barrels" differed from responses to the original item. Indeed one respondent covered the whole range by "disagreeing very much" for the dishes item and agreeing just as strongly for "children." Faced with the problem of a compromise response to the original item, he had decided to "agree pretty much."

That the single-barrelled items are more reliable than the original form is strongly suggested by the case of one couple. Their responses to the original form differed sharply with scores of five points for the husband and two for the wife. At first glance, these contrasting scores appear to indicate that he is rather authoritarian in his family role expectations whereas she is comparatively equalitarian. Indeed this couple would be poorly matched if he were disagreeing "pretty much" with the same idea that his wife agrees to. However, their "developmental scores" on the writer's Traditional-Developmental Scale show close husband-wife similarity at 10 and 11 points respectively (2). This similarity is confirmed when the couple are able to respond independently to the separate factors in Item E. Under these circumstances both partners score five regarding dishwashing and agree on the extreme one-point position regarding the care of the children. Apparently the husband-wife discrepancy in scores for Item E reflects the awkward problem involved in knowing how to weight one's answer to a double-barrelled question.

For the 24 respondents as a whole, mean scores on the two sub-items differed substantially from each other and from the original phrasing. With respect to children, the mean was close to the equalitarian extreme at 1.58. Scores regarding dishwashing averaged almost two notches higher at 3.50. For the original item, the mean fell in between at 2.29. These markedly divergent means show that the item as originally formulated requires respondents to give what may be a misleading over-all reaction to relatively incompatible ideas.

The similar ambiguity of two other items in the A-E Scale may be indicated by inserting parenthetical lettering: (Item A) "Human nature being what it is, there may always be (*a*) war and (*b*) conflict." (Item G) (*a*) "People sometimes say (*b*) an insult to your honor should not be forgotten."

Differential reactions to the separate elements in such items suggest the confusion created when a single response is expected to complex stimulation. Whereas in projective testing such confusion serves a useful purpose, in attitude measurement the desired precision and clarity can be achieved only by avoiding ambiguous phraseology.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.
2. BLOOD, R. O., JR. Developmental and traditional childrearing philosophies and their family situational consequences. Unpublished doctor's dissertation, Univer. of North Carolina, 1952.
3. SANFORD, F. H., & OLDER, H. J. *A short authoritarian-equalitarian scale.* Philadelphia: Institute for Research in Human Relations, June 1950.

Received January 25, 1954.

THE SELECTION OF STIMULUS WORDS FOR VALUE: DURATION THRESHOLD EXPERIMENTS

DAVID ROSENTHAL

Johns Hopkins University, School of Medicine

IN A recent review of the literature, Eriksen (1) concluded that the data clearly supported the concept of "perceptual defense" and he indicated some directions for future research in this area.

Our aim here is to raise two points which seem not to have received sufficient emphasis in previous reports, and which warrant consideration in further research. The first involves the selection of stimulus words in those experiments which purport to show a relationship between the duration of recognition thresholds and the values which an individual holds. We do not refer to the use of socially objectionable words made by McGinnies (4) and others. The problems here have already been discussed at length, and reviewed again by Eriksen (1).

We are concerned rather with the problem of selecting words without first ascertaining their appropriateness. Postman, Bruner, and McGinnies (5) and Postman and Schneider (6) used as a criterion for selection the agreement of judges that the stimulus words were related to given values, while Solomon and Howes (7) took words directly from the Allport-Vernon Scale of Values, and used them out of context. Both procedures are open to question.

The second point concerns some lack of clarity in the use of the term "values." The distinction between value as "interest" and value as "preference" warrants further emphasis. A person may have a high preference for "God" and a low preference for "devil" yet be equally interested in both.

Postman, Bruner, and McGinnies stated that they were using the term value in the sense of "personal interest," yet they defined "contravaluant responses" as words which "were opposite in meaning to the stimulus word or served to derogate it" (5, p. 149). Here the implication seems to be one of preference. Haigh and Fiske indeed speak of the PBM findings as being concerned with "value preferences" (2, p. 394).

To indicate the degree and kinds of error that can occur in the selection of stimulus words, the author used a modification of the Haigh and Fiske procedure of ranking all stimulus words by value preference. The high frequency word set of Solomon and Howes (7, p. 260) and both the high and low frequency word sets of Postman and Schneider (6, p. 278) were reproduced by rows, but the order of the words in each row was changed to achieve a randomized sequence. Ten subjects were asked to rank the words in each row from 1 to 6 according to their preference.

Kendall's W (3), called the coefficient of concordance, was used to measure the degree of homogeneity of ranks for each word set. This measure ranges from 0 to +1, zero indicating no agreement in a set of m rankings of n objects, +1 indicating perfect agreement.

In selecting stimulus words, we aim at a high order of homogeneity, meaning that all words representing a given category tend to be given similar ranks. If the ranks are scattered, we may question whether the words selected really belong in the same category.

For half our subjects, the homogeneity of the Solomon and Howes high frequency word set did not even reach a value significantly greater than zero, the mean $W = .43$. On the Postman and Schneider set of frequent words, statistical significance was obtained in 3 out of 10 cases, and on the set of infrequent words, statistical significance was obtained in 4 out of 10 cases. Thus, we find a great deal of error in the grouping of these words under their respective categories, when we consider values to be concerned with "preferences."

Some qualitative observations shed further light on the problem. Some words seemed to serve as stereotypes. For example, the modal rank of "liberties" was 1, whereas the modal rank of "politician" was 6, although both were presumed to indicate political values. "Truth" had a high modal rank, "physics" a low rank, yet both were supposed to represent theoretical values. Words like "truth," "liberties," "loving," "friendly," and "kindness" seemed to be so general in their implications and so highly, positively charged that they seemed to transcend classification in any one category. Other words like "politician," "status," and "dominant" aroused negative connotations and received low modal ranks.

Some words were ambiguous. Several subjects thought of "conception" not at all as a scientific word, but rather in connection with sex and childbirth. "Heavenly" was thought by one subject to be a typical fatuous comment by a modern teen-age girl, rather than a word bearing on the truly religious and spiritual. "Confession" was not always thought of in its religious sense, or "efficiency" in its economic sense. Other examples could be cited.

Words were not of comparable specificity or generality. A specific object like "automobile" or "painter" tended to receive low ranks as compared to more general terms in the same category like "efficiency" or "graceful."

It is possible that if our subjects had been asked to rank the words according to interest rather than preference, greater homogeneity of ranks would have been achieved. However, Haigh and Fiske (2) found that when words were ranked by value preference their association with duration thresholds was sharper and clearer than when the words were used presumably as indicators of interest. One might deduce from this that interests are related to duration thresholds and to perceptual defense mechanisms only insofar as they reflect an individual's preferences.

In future experiments greater care in the selection of stimulus words, possibly by using the method suggested here, would eliminate much error variance and permit experimental relationships to be brought out more clearly. Greater elucidation of the psychological distinctions between value as "interest" or "preference" should help to clarify further the concept of perceptual defense.

REFERENCES

1. ERIKSEN, C. W. The case for perceptual defense. *Psychol. Rev.*, 1954, **61**, 175–182.
2. HAIGH, G. V., & FISKE, D. W. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1952, **47**, 394–398.
3. KENDALL, M. G. *Rank correlation methods*. London: Charles Griffin, 1948.
4. MCGINNIES, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
5. POSTMAN, L., BRUNER, J. S., & MCGINNIES, E. Personal values as selective factors in perception. *J. abnorm. soc. Psychol.*, 1948, **43**, 142–155.
6. POSTMAN, L., & SCHNEIDER, B. H. Personal values, visual recognition, and recall. *Psychol. Rev.*, 1951, **58**, 271–284.
7. SOLOMON, R. L., & HOWES, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58**, 256–270.

Received February 23, 1954.

RECIDIVISM AMONG GIRLS

WILLIAM W. WATTENBERG AND FRANK SAUNDERS

Wayne University

THIS is an empirical study of factors associated with repeating among girls contacted on complaint by the Youth Bureau of the Detroit Police Department in 1952. It was hoped to find answers to three interrelated questions: (*a*) What items of information distinguish girls with previous police records from those who are in trouble for the first time? (*b*) How does the list of discriminating factors for girls differ from that for boys? (*c*) What does this comparison imply as to the processes whereby girls, as contrasted with boys, become delinquent?

The populations employed consist of all boys and girls under seventeen years of age against whom complaints were made. When boys or girls are taken into custody in Detroit, they are promply turned over to designated officers attached to the Youth Bureau. To deal with boys, there is a small detachment of Youth Bureau officers in each precinct. For girls, a group of policewomen known as the Girls' Behavior Section is attached to the central office of the Youth Bureau.[1]

For each youngster, the appropriate officer or policewoman makes an investigation which routinely provides for an interview with the youngster and an interview with the parents. In most cases the home is visited. On the basis of all the information secured a "history sheet" is completed. In 1952, this called for facts or ratings on sixty items relating to home conditions, school, neighborhood situations, and the youngsters' activities. The data thus recorded were coded and punched on IBM cards. These cards were the basis of the present study.[2]

The procedures used in this study were simple. The cards for all girls contacted on complaint in 1952 were divided into two groups in accordance with whether or not the files of the Girls' Behavior Section showed evidence of the girl having been the subject of a police contact more than once. For every item of information a tabulation was prepared comparing the repeaters with the nonrepeaters. Each such tabulation was submitted to the chi-square test to determine the degree of confidence with which the null hypothesis might be rejected. The same operations were performed on the records of the boys.

For 1952, there were records for a total of 1,082 girls, of whom 427 were repeaters and 655 were one-time offenders. Of the 60 tabulations made possible by the data, the null hypothesis could be rejected at the .01 level of confidence for 21, and at the .05 level for seven more. (By chance, no more than three tabulations might be expected to show these levels.)

Table 1 contains the major facts on the 28 items for which the null hypothesis could be rejected at the .05 level. In the first column of the table, the subject matter of the item is described. The second column lists those catergories for each item in which the percentage was high for repeaters; the third column gives the value of p.

It will be noted that these discriminating items tend to fall into a few major clusters. Of these, one of the most striking gives evidence that the girl repeaters were more in conflict with their homes than the nonrepeaters. The girls' mothers appeared to play the more significant role in the situation.

School also appeared to be an area of tension for the girl repeaters. More of them had quit school. Aside from the poor relations with school classmates, the girl repeaters seemed to enjoy good peer-group relationships more than the nonrepeaters.

None of the direct socioeconomic indices distinguish reliably between repeaters and nonrepeaters.

As to the girls themselves the repeaters were older, more mature, and more likely to be well kept.

It is only by comparing these trends for girls with the results of a similar study for boys that we can obtain clues as to the extent to which the above findings represent a typically feminine pattern rather than one which applies to all youngsters.

There was a difference in the categories which contributed to this result for three items which had been significant for both girls and boys. For boys the repeaters were more often charged with

[1] Cases originating as missing persons reports or in arrests of adult offenders are usually handled directly by the Women's Division at police headquarters and often do not reach the Youth Bureau.

[2] For assistance in connection with the study, the authors wish to express appreciation to the following police officials: Senior Inspector Sanford Shoults, in charge of the Youth Bureau; Inspector Ralph Baker, in command of the day-to-day activities of the Bureau; Lieutenant Francis Davey, who worked out the details of the recording and coding system; Sergeant Elizabeth Grindle, in charge of the Girls' Behavior Section; and Sergeant Mervin Roberts, who handled the IBM operations for the Youth Bureau while the study was in progress.

TABLE 1

ITEMS SHOWING ASSOCIATION WITH REPEATING AMONG GIRLS AT THE 5 PER CENT LEVEL OF CONFIDENCE

Item	Categories Positively Linked to Repeating	Categories Positively Linked to Nonrepeating	$p <$
Expressed attitude toward home	Dislikes; indifferent	Likes	.01
Attitude expressed toward mother	Dislikes; makes no comment	Likes	.01
Attitude expressed toward both parents	Dislikes mother	Likes both	.05
Form in which parents give money	None given	Money given on request	.01
Parents' attitude toward police officers	Indifferent	Cooperative	.01
Hours when mother is at home	None	Evenings; all the time	.05
Parents' participation in girls' activities	"Seldom"	"Occasional"	.05
With whom girl is living	Stepfather and own mother; boarding parents; father alone	Both own parents; stepmother and own father	.01
Offenses charged against girls at 13 or 14 years of age	Incorrigibility; truant from home	Larceny; sex offenses	.05
Offenses charged against girls at 15 or 16 years of age	Incorrigibility; truant from home	Larceny; sex offenses	.01
Relations with other youth	Member of a regular group	"Lone wolf"	.01
Group activities	"Rambunctious"; organized theft	Orderly	.01
Grade in school	Not attending; in special classes	Below eighth or in eleventh	.01
Estimated "mentality"	"Below average"	"Above average"	.01
Expressed attitude toward school	Dislikes; indifferent	Likes	.01
Attitude toward teachers	Contempt; no comment	Friendly	.01
Relations with classmates	Ignores; quarrels; shuns	Friendly	.01
School marks	Poor; failing	Average; above average	.01
Expressed attitude toward adult neighbors	Indifferent	Friendly	.01
Age	14–15	12 or under	.01
Sexual development	Early adolescence	Preadolescence	.01
Girls' appearance	Fastidious; overdressed	Slovenly	.05
Sports and hobbies	None; dancing	Games	.05
Size of home	2 rooms per person	1.0 to 1.5 rooms per person	.02
Condition of dwelling unit	Needs repairs	Average	.01
Condition of dwelling interior	Unduly meticulous; slovenly; all others	Average	.01
Officer's judgment of girls' attitude toward officer	Evasive; "antisocial"	"Honest"	.01
"Type" of youth	Quarrelsome; mischievous	Studious; "normal"	.01

offenses against property rather than actions directly concerned with the home. Also, the relative neglect by parents was more extreme among boy repeaters. Interestingly, overcrowded homes were more closely linked with boys' repeating; whereas girl offenders from spacious homes were more likely to be repeaters.

An analysis of the items which had been statistically significant for boys, although inconclusive for girls, threw additional light on the above finding. One outstanding cluster of such items related to socioeconomic variables. For these the trends were paradoxical. While they were more closely linked to boys' repeating, yet the conditions were worse for girls than for boys.

This same paradoxical situation prevailed on items designed to reflect family tensions. Showing this trend were the following additional items bearing on family situations:

1. Presence of father in the home
2. Expressed antipathy to father
3. Reported intensity of family quarreling
4. Expressed antipathy to stepfather
5. Youth living with relatives other than parents
6. Youth living with mother alone
7. Parents seeking to alibi selves in interviews with police

A number of other items, not related to the above clusters, distinguished boy repeaters but not girl repeaters. These were:

1. Youngster rated "small for age"
2. Church attendance reported as irregular (less than once a week)
3. Preference for going to shows as leisure-time pursuit
4. Nonparticipation in organized youth groups
5. Youngster does no chores around the home
6. Youngster had part-time work doing odd jobs, rather than having a newspaper route (Very few girls held part-time employment)

SUMMARY

This study was based on information gathered by the Youth Bureau of the Detroit Police Department on all girls reported to its Girls' Behavior Section in 1952. The records of 427 repeaters were compared with those of 655 nonrepeaters. Of the items associated with repeating, the largest clusters relate to friction with parents and with schools. As compared to a similar study of boys, repeating among girls was linked to a more narrow range of factors. Girls more than boys were affected by present relationships within the home, particularly those involving their mothers.

Received March 20, 1954.

PSYCHOTHERAPY AND PERSONALITY CHANGE. Edited by Carl R. Rogers and Rosalind F. Dymond. Chicago: Univer. of Chicago Press, 1954. Pp. x + 447. Price $6.00.

This volume, with twelve authors, edited by Carl Rogers and Rosalind Dymond, constitutes essentially a progress report on the large-scale research program in psychotherapy which has been under way at the Counseling Center of the University of Chicago for more than four years. It presents findings to date on outcomes and process of the client-centered or nondirective approach to psychotherapy. The research is motivated by the desire to test four specific hypotheses which are central to client-centered therapy in regard to the capacity of the client who seeks help, the function of the therapist, and the process of therapy. The empirical support or disproof of each hypothesis, according to Rogers, tends to confirm or to cast doubt upon a whole body of theory in the area of personality, learning, and perception. The research program has grown out of the context of earlier research in which Rogers and other members of the group of authors whose work appears in this volume have been engaged.

The book is divided into four parts, Part I presenting the context and the basic design of the research, Part II describing a series of separate investigations bearing on the various hypotheses, and Part III giving what the authors call an "objective analysis" of the therapy of two single cases, one "successful" and the other "unsuccessful," and attempting to relate each of them to the findings of the major studies reported in Part II. Part IV is called an "Overview" and poses many stiff questions for the future. The Introduction is marked by both pride and modesty. This is a pioneering effort and is hailed as such with satisfaction, but the difficulties and imperfections both of method and of investigators are not minimized. The orderliness of the Introduction permits the reader to see at a glance which portions of the volume are of primary interest to him.

Of the 11 studies of which the findings are given in Part II, most are concerned with the investigation of some aspect of the third central hypothesis, namely, that in the acceptant, understanding atmosphere provided by client-centered therapy, the individual will change in his perception of self, will become more understanding and accepting of self and others, more creative, adaptive, autonomous, tolerant, mature, and less defensive. This is a broad and inclusive hypothesis, and most of the studies have some bearing upon it. The *Q* technique furnishes the most used and most useful of the methods employed, but the instruments used include also a variety of psychological tests previously developed by others and some developed for the purpose, including rating scales. The various studies report decreasing discrepancies between concepts of the "self" and of the "ideal," and between concepts of the "self" and of the "ordinary person" in the course of therapy. The relationship of such decreases to "adjustment" level and to increase in "self-awareness" during "successful" therapy is also studied, and the effect of therapy on attitudes of acceptance and respect for others and on maturity as measured by the Willoughby Emotional Maturity Scale comes in for investigation. All of these fall squarely within the area of the third hypothesis. Five other studies, though concerned with measurement of the same variables, devote primary attention to methods of measurement or prediction of change (such as counselor judgment and ratings based on the TAT), and to the relation of change in therapy to such factors as ethnocentrism and motivation.

The authors insist that they will avoid value judgments in describing outcomes of therapy because there is no general agreement about the nature of adjustment or maladjustment. But this is a hard promise to keep, and the reports are full of such words as "improvement," "favorable changes," "better mental health," and so on, which are seldom placed within quotation marks. The circumlocutions required to describe outcomes in terms of changes in relation to the concept of the self or of others, or of flexibility in perception or in ease of learning are still too clumsy.

One of the novel and heartening elements in the research design is the plan for two control groups, the "equivalent" control group which had no therapy but regular testing, and the "own controls" or "wait" group. Half of the clients who agreed to participate in the research agreed also to defer beginning therapy for 60 days. Since test batteries were administered before and after the 60-day period, it was possible to compare changes in the waiting period with those taking place during the therapy period. Subsequent variations in behavior of this "wait" group permitted observations on results of differences in motivation for therapy. This aspect of the design is a first attempt to approach and explore the little-understood phenomenon of "spontaneous recovery." The findings suggest that some of the less disturbed clients who during the 60-day period had made positive gains on many of the measures used never really became involved in therapy.

Another unusual aspect of the research is not in design but in method. Probably more than in any other major investigation, these studies utilize as

their basic data the individual's own phenomenological field. From it his private frame of reference is inferred. By the use of the *Q* sort to obtain pictures of the self, the self-ideal, the other person, of degrees of awareness of self, and so on, it is possible to derive the client's view of certain aspects of his private world. A certain lawfulness and order are discovered within these phenomenological data, both in relation to external observations and in relation to the individual's internal scheme. Rogers, though he recognizes that there are dilemmas involved in the method, nevertheless regards it as by far the most rewarding that he has used and the most fruitful in leading to new and deeper hypotheses.

The "research analyses" of the therapy of two cases written by Rogers are the most intensive and complete yet presented in the literature. Both are attempts to obtain solid, objective evidence as to what goes on in the course of therapy, and both give extensive excerpts from the actual interviews in order to communicate the flavor of the sessions. The findings on each client draw upon the investigations within six different research projects. The "failure" case is a somewhat confused part of the story, especially since the counselor rating on outcome of therapy for this client was well above the mid-point on a 9-point scale. Failure to change in therapy is not included among the situations for which specific hypotheses were developed at the outset, and the occurrence is viewed with some puzzlement. Rogers suggests two hypotheses regarding "failures" in therapy as a result of analysis of this case—one, that "long" therapy is more likely to be helpful than "short" (a position to which the nondirective group has been moving with more and more certainty over the years), and two, that perceived safety in the therapeutic relationship is essential if disturbed clients are to continue in therapy. Even these hypotheses, however, which Rogers proposes for future testing, do not seem to get at the heart of the matter or to explain why the capacity for productive self-reorganization becomes released for some and not for others in the accepting climate of psychotherapy.

This is an impressive book for many reasons. Impressive are the courage, honesty, and industry shown in approaching this almost impossible task of research in a field so full of obstacles. It is no accident that this is in fact a "pioneering" program and that many generations of psychotherapists have shrunk from so formidable an undertaking. It is a task that demands not only the willingness and the candor to lay all the facts on the line. It demands also a capacity for hard work and tedium, for spending, for example, more than five hundred man-hours of effort simply in collecting and transcribing the data from one client and his matched control. It demands further a kind of group devotion to a central body of theory and its related practice, and resourcefulness in developing testable hypotheses arising from the theory. The followers of the client-centered approach around Carl Rogers have been able to develop this group enterprise. Certainly it is high time for this kind of investigation, and even now no other school of therapists is zealous or stalwart enough to attempt it.

In the relation of the findings to general psychological theory one feels a disappointment in this volume. Perhaps this is in part because we want too much, because we want too clear and orderly and rich a body of theory to emerge. Perhaps it is in part because our hopes are raised too high in the early discussion of theory in the book. Certainly the studies reported here have little bearing on the field of learning theory or on that growing up around the process of perception, as the preview leads us to expect. That aspect of theory which is explored here and which finds support in the findings of the various studies is chiefly the "self-theory," the self and the self-concept in relation to the process of psychotherapy and the therapeutic relationship. Even in this area of theory, however, there is a fragmentary quality and the reader is bedevilled by many of the doubts which also trouble Rogers, by the conflict among "perceptual vantage points," and by the relation of these varying perceptual systems to an integrated personality theory. Perhaps in this complex field for a long time to come we must content ourselves with bits and pieces, and only much later expect to reach the well-ordered and integrated structure of theory that we desire.

RUTH TOLMAN

Pasadena, California

PSYCHOANALYTIC INTERPRETATION IN RORSCHACH TESTING. By Roy Schafer. New York: Grune & Stratton, 1954. Pp. xiv + 446. Price $8.75.

This is probably the most significant and least advertised book in clinical psychology to appear in many a year. It is also one of the very few books on psychological testing addressed to the advanced worker. Many have been the books on Rorschach for novitiates, detailing how to administer and score the test; few have been the books on how also to interpret the results; and almost nonexistent have been the books dedicated to content analysis of the Rorschach. This is such a book. Additionally it is, to the best of this reviewer's knowledge, the first book that explicitly attempts to root the entire test firmly in a given personality theory, in this case the neo-analytic ego theory of the Freudian school.

This book officially initiates a new trend in Rorschach interpretation, a trend that is heavily

indebted to E. Schachtel. Schachtel was one of the first to advocate a method of viewing the Rorschach that avoided the moribund approach of "scores" and "signs." Bringing to the Rorschach astute clinical insights, he substituted instead a dynamic interpretation of the test based largely upon the theoretical contributions of Harry Stack Sullivan and Erich Fromm. Schachtel demonstrated how the test could be utilized to ascertain the individual's attitude toward the test, the examiner, the entire test situation, and even to his own responses toward the inkblots, and how, from this information, the personality structure of the individual could be deduced. Taking his cue from Schachtel, but attempting to go beyond it, Schafer stresses content interpretation to the extent of eliminating most of the usual considerations of scores, percentages, and the like—not that such scores are completely unimportant, Schafer is careful to point out, but because he wishes to demonstrate how much may be achieved without them. Clearly, his book starts, not where other books leave off, but where other books have yet to probe.

Feeling that what one obtains from any test depends upon what one is looking for, Schafer firmly advocates a theoretical frame of reference in interpreting the Rorschach. The theoretical frame of reference to which he subscribes is not the classical Freudian one, but rather a more modern "ego" psychology that is greatly indebted to the work of the Rapaport group, of which Schafer is a member. This group has attempted to revise and to integrate into the original Freudian theory new insights and findings and then attempted to apply these to the domain of the clinical psychologist, primarily to test interpretation.

The approach centers about four central areas: (*a*) the test situation and relationship as they bear upon the perception, hence the responses, of the client. As indicated, this area is most directly indebted to Schachtel; (*b*) levels of psychic functioning, which rests upon Freud's concepts of primary and secondary thought processes and Kris' concept of "regression in the service of the ego." The central thesis is that the client is constantly undergoing shifts in the amount of intellectual control that he exercises during the test, much as any individual will shift, at various times during the day, from sharp, problem-solving thinking to "free" fantasy in the form of daydreams; (*c*) patterns of defense, with particular attention paid to the defenses of denial, reaction formation, projection, and repression. Here, the debt to Freud and to Anna Freud is obvious. This section constitutes the heart of the applied aspects of the book, the author carefully demonstrating how an evaluation of both the type of defence and the efficacy of the defense can be gleaned from the content of the responses to the Rorschach; (*d*) Ego identity, particularly as postulated by E. H. Erickson, although many may feel that at least a part of this debt is to Gordon Allport and to Carl Rogers. Schafer utilizes this concept to demonstrate how a knowledge of the person's self concept, his age, cultural background, vocation, and similar areas of importance to ego identity can be utilized to change the shading of an interpretation in order to make it more highly individualistic and unique to the particular individual being tested.

To state the origins and indebtedness of the author to other workers in the four areas cited is not to take credit from Schafer. He himself is more than magnanimous in acknowledging such debts, but he too must be given credit for the careful revision and adaptation of these concepts to fit the Rorschach situation as he perceives it.

It might be thought that any test analysis dealing with Freudian theory would lend itself to what Schafer calls "wild" analysis and symbolic interpretation. It is a great relief to find that Schafer, unlike many, pleads fervently against such an approach and offers, in addition to his plea, a list of criteria for avoiding this type of analysis. Constantly reiterated is the admonition that one must remain close to his data when interpreting the test. It is not enough, for instance, when one finds a passive attitude to be basic in a client, to jump directly into theory and to state that this is an "oral receptive individual." Evidence for both the receptivity and the orality must be present in the test results themselves; without evidence for orality Schafer would propose that the examiner write only that "this is a receptive person," although one might, of course, want to postulate that probably an oral fixation was involved. In his chapter, "Criteria for Judging the Adequacy of Interpretations," specific safeguards are proposed which, if followed by those in need of them, will inevitably lead to more conservative, less gratuitous content interpretation.

Similarly rewarding are Schafer's suggestions for organizing interpretations into hierarchies and for the avoidance of the usual clichés and vapid spinning out of "tendencies," a catchall phrase used by many clinicians to indicate that there will follow all those disparate segments of interpretation that the examiner was unable to organize into meaningful patterns. Schafer demonstrates precisely how such organization can be effected and how it can profitably replace the usual string of unconnected adjectives that constitute the "tendencies."

One could go through page after page of this book pointing out the vein of gold within each. This book represents a thoughtful, critical work by an extremely gifted diagnostician and clinical

psychologist. It is practically inconceivable that a clinician could go through any ten pages of the book and not find something from which he could greatly profit.

MARTIN KAPLAN

Peoria State Hospital

TRANSFERENCE. ITS MEANING AND FUNCTION IN PSYCHOANALYTIC THERAPY. By Benjamin Wolstein. Introduction by Clara Thompson. New York: Grune & Stratton, 1954. Pp. xiii + 206. Price $5.00.

The analysis of transference is generally considered to be the essential technical characteristic of psychoanalysis. In fact, the analysis and interpretation of the irrational components of the patient's view of the therapist are widely accepted as the defining operation of psychoanalysis together with the use of free association to recover repressed childhood memories and attitudes. Surprisingly enough, despite this fact, relatively little attention has been paid to transference in the general psychological literature or even in the works of learning and perception theorists who have sought to translate analytic concepts into broader behavioral schemes. With the developments in ego psychology and the study of character defenses that have marked the past fifteen years or so there has been gradually more concern with the problem of transference and countertransference, as evidenced by an increased sprinkling of articles and symposia in the professional journals. Wolstein's book has appeared at an important time, and in future years it may well be regarded as a pioneering effort at presenting an integrated critical analysis of the concept of transference in psychoanalysis.

After a lucid introduction by Clara Thompson, which summarizes historical trends in the concept of transference, the author clarifies the objectives of the monograph. Essentially this is a critical study, by no means pretending to be a comprehensive survey of the field. There is an initial attempt to orient the reader toward certain basic scientific postulates from which the conception of transference is derived. The philosophic emphasis is somewhat unusual for a book on analytic technique. Yet in view of the general objective of clarifying the concept of transference, the stimulating discussion of operational procedures in scientific theory construction is quite valuable. Wolstein attempts to orient the reader toward the essential philosophical assumptions which underlie the theoretical constructs of Freud, who worked essentially in the atmosphere of a mechanistic, monistic, reductive scientific world. A contrast between these assumptions and the more recent implications of scientific philosophy, emphasized in the writings of Dewey, James, Mead, and Pierce, is drawn to demonstrate the need for a view of transference that goes beyond a single factor theory of personality or a reductionistic and historical concept of the etiology of current behavior patterns. Subsequent chapters trace in detail the origins of Freud's concept of transference. The essential point made by the author is that Freud's concept of transference, stemming as it did from the early use of hypnotic suggestion and catharsis, never quite lost this quality of the employment of the irrational authority of the therapist as a curative technique. Transference, as Wolstein teases it out of various writings of Freud over his lifetime, becomes in its positive form a tool of the analyst, the equivalent of the hypnotic influence which is geared toward recovery of lost memories. In essence, argues Wolstein, Freud's concept of transference involves merely the more effective use of suggestion to influence not so much the content of the patient's symptomatology as the character resistances which are obstacles in the analysis to the recall of traumatic childhood memories. Thus, as Freud describes the process, the transference reactions which develop in analysis are usually ambivalent; but when *positive*, i.e., characterized by love attachment or ascription of unusual power to the therapist because of unconscious dependency needs, they are the chief ally of the analyst. Certainly a basic rule of orthodox analytic technique has been to analyze immediately all negative transference and refrain from analysis of positive transference. Presumably under the sway of the positive attachment the patient is more likely to strive to do the analytic work well and to recover the repressed childhood fantasies. The specific point at issue in this monograph is whether such a view of transference is tenable in the light of current philosophical and psychotherapeutic sophistication. Surveying Freud's own writings and discussing in detail the case of Dora, Wolstein argues that Freud's concept is psychologically unsound, authoritarian, and ineffective in the long run in leading to personality change.

In contrast to Freud's approach, Wolstein presents a point of view stemming in part from his own experience and that of a number of analysts associated with the theoretical reconstruction of Freud's work formulated by Harry Stack Sullivan, Erich Fromm, Clara Thompson, and, to some extent, Karen Horney. This *genetic-functional* point of view emphasizes the communication process and draws chiefly on the more operational concepts of unconsciousness exemplified in Sullivan's contrast between *parataxic* and *syntaxic* modes of relating and thinking. The goal of analysis becomes exposure in the here-and-now medium of patient-therapist interaction of *all* of the distorted components of the patient's reaction.

The therapeutic process becomes not so much recall of the repressed as examination of the distortions of the present or future aims and perceptions of the analysand. The past is not to be ignored, but recall of childhood becomes a means of providing clear perspective rather than the cathartic goal which it presumably is in more orthodox analyses. In subsequent chapters Wolstein elaborates on this point of view and attempts an operational formulation of transference and countertransference. The additional key to the meaning of transference distortions, as a special case of *parataxis*, is the occurrence of anxiety in the course of communication. Wolstein concludes with a challenging discussion of the implications of the reformulated concept of transference for value theory and metapsychology.

So sketchy an outline can scarcely do justice to the richness and subtlety of the author's argument. It should be apparent to the reader, however, that this is indeed an unusual and provocative work. What is most striking about it is perhaps the nonclinical orientation. Transference emerges in these pages as the manifestation in the analytic situation of a pervasive distortion of perception and value that impedes productive living and effective social development. Analysis becomes in this book a new type of creative experience for people with difficulties in living rather than a form of treatment for mental illnesses of various types. This is a perspective that may well appeal to many readers otherwise not concerned with psychotherapeutic practice. It may also serve as a useful corrective for clinicians too deeply immersed in technique and psychopathology to recognize some of the broader implications of their work. The book has a self-searching quality that stems from a deep understanding of the need for psychotherapists to examine not only their own feelings in relation to a given patient but their general orientation, which must inevitably be communicated to patients in the course of their work.

Certain limitations of this monograph, imposed in part by the author's objectives and in part by the approach, may be mentioned briefly. In restricting his presentation to a critique of Freud's own writings, Wolstein necessarily omits the developments and increased richness in the view of transference within orthodox analytic literature. Occasionally one feels that in his somewhat polemical treatment (characterized by excessive repetition of his main point about the hypnotic roots of Freud's concept of transference) the author is fighting a fleeting ghost from the past. Certainly in the recent writings of Alexander, Balint, Macalpine, Little, and La Gache the issue of the analyst as a mirror or the toleration of positive transference is seriously questioned by analytic writers. Consideration of the wider body of literature in the field might have been more helpful to the reader in evaluating Wolstein's point of view. Similarly, despite an unusually fluent and gifted prose style, the author's failure to include case materials and examples seriously interferes with the communication of his message. In a book that deals so brilliantly with the understanding of human personality one is a bit dismayed to find so little reference to actual people. This overly abstract presentation is perhaps merely a stylistic impediment, but it may seriously hinder evaluation and appreciation of the author's point of view and his criticisms of Freud's concepts. Finally, one looks in vain for some documentation of the repeated assertion that the *genetic-functional* point of view is therapeutically more valuable. This is perhaps too much to expect from an essentially critical thesis, and it may be that by attempting to trace out the meaning of transference in so clear a fashion Dr. Wolstein's contribution will lie in stimulating others to direct empirical investigation. This is undoubtedly a difficult, somewhat overwritten, and occasionally exasperating book, but it is a brilliant one that raises important issues, faces up to them squarely, and goes to the core of many problems in current psychotherapeutic theory and technique. Practising therapists, personality theorists, social scientists, and advanced graduate students will find it stimulating and provocative of serious self-searching.

JEROME L. SINGER
Franklin D. Roosevelt VA Hospital
Montrose, New York

THE CLINICAL INTERACTION: WITH SPECIAL REFERENCE TO THE RORSCHACH. By Seymour B. Sarason. New York: Harper & Brothers, 1954. Pp. x + 425. Price $5.00.

The argument of this discursive but always piquant book runs about as follows: Rorschachers too often jump from test scores to a prediction of behavior. There ought to be an intervening variable to provide some rationale for such leaps. Besides, recent empirical attempts to validate the Rorschach show that the equalities traditionally assumed between scores and behavior are not always real. Better interpretations might be obtained by using as an intervening rationale "the clinical interaction" that goes on during the test. From this construct, especially as it leads to an understanding of how the subject was perceiving the test, one can make inferences about his probable behavior that will be both more logical and more valid. Introspective accounts of the analysis of six protocols are offered to show how this approach works.

The reader finds obvious flaws in each step of the argument. The whole thesis depends upon

setting up straw men. Even reasonably trained Rorschachers aren't so bad as Sarason assumes the group as a whole to be; when the reader discovers him quoting Klopfer on the side of the iconoclasts, he begins to wonder who, in Sarason's eyes, represents standard usage. The empirical studies mentioned are interesting—and the book is valuable as a compilation of recent validation experiments—but most of them measure the usefulness of the Rorschach against some paper and pencil test, with shocking unconcern for the known flimsiness of such instruments. One has the feeling that most of these studies should be read the other way around: the Rorschachs demonstrated that So-and-So's list of a priori questions did not in fact reveal the trait So-and-So assumed he had measured. Sarason suggests revisions of interpretation on the strength of these studies and of recent factor analyses of Rorschach scores. Many of his suggestions make sense, but the reader wonders if they do not do so less because of the thin evidence on which they are ostensibly based and more because Sarason carries to its end the original logic of Rorschach: the polarity between responses arising from within the person and responses arising from external stimulation by the inkblot.

The choice of the clinical interaction as an intervening construct is also not new. Even when learning the use of nonprojective tests, most clinical students have been taught to watch what transpires between themselves and the subject. When this rule is phrased in terms of being aware of how the subject has structured the task, it becomes a cliché. For instance, no one should find it a new idea that high $F+\%$ occurs when the subject has decided it is his job to be accurate. Yet by carefully tracing out the interpretive implications of the subject's self-definition of the inkblot task, Sarason does arrive at some stimulating corollaries. The reader may or may not, for example, feel that C' is merely a form of C; the idea is at any rate interesting. (In this reviewer's experience, it seems valid for very bright normal people but not for pathological cases.)

The step Sarason does not take is to see that if the test is a behavior sample collected under and understood in terms of current conditions, it is therefore also a sample of the premises that the individual brings to all situations, whatever the conditions. At a second degree of abstraction, the clinical interaction tells us about the lifelong premises, the treasured self-consistencies, with which the subject has learned to meet the world. A lifetime of overdetermination brings about and may be read from the clinical interaction seen during a Rorschach session, but at no point does Sarason take his own premises or the Rorschach that seriously.

Reading Sarason's introspective accounts of his analysis of six Rorschachs, the reader who is himself a Rorschacher will find little evidence that what Sarason does is different from standard usage. It would seem that the incendiary tone occasionally adopted in the theoretical section of the book was inappropriate, however stimulating it may have been. That more such examples of the interpretative process should see the light of print is very desirable, however, and their inclusion alone might justify calling attention to this book.

On the whole, *The Clinical Interaction* is lively reading. One finds pleasure in disagreeing and pleasure in agreeing—on alternate pages!

CHARLES MCARTHUR
Department of Hygiene
Harvard University

The reviewer says, (It is) "the first really serious attempt to work out empirically the full consequences of Kurt Lewin's approach."

MIDWEST AND ITS CHILDREN

THE PSYCHOLOGICAL ECOLOGY OF AN AMERICAN TOWN

By **ROGER G. BARKER**, *Professor of Psychology, University of Kansas* and
HERBERT F. WRIGHT, *Professor of Psychology, University of Kansas*

Now for the first time accurate information is available on what actually happens to children and how they behave as they live their lives in one town.

This new book describes in plain language and in concrete detail the everyday behavior of the children of the town of "Midwest, U.S.A." and the ordinary circumstances of their lives as seen by skilled observers over a six-year period. It includes description, in terms of theories and concepts, of the behavior and living conditions of these children. Comparative data are included upon children in a school for crippled children situated in a neighboring city.

Write for full details.

Evanston
Illinois

Row, Peterson and Company

White Plains
New York

Dynamic and Abnormal Psychology

W. S. TAYLOR, Smith College

"A very fresh and original presentation."—G. W. Williams, Rutgers University

"It is refreshing to see emphasis on phenomena instead of psychiatric syndromes."
—Mortimer Garrison, Jr., University of Pennsylvania

"An excellent discussion of the basic psychological phenomena necessary for the understanding of both normal and abnormal behavior."—T. E. Hannum, Iowa State College

Social Psychology • An Interdisciplinary Approach

HUBERT BONNER, Columbia University

"The interdisciplinary aspect of this book answers a longfelt need. It is a real contribution."
—Gordon W. Blackwell, University of North Carolina

"The book is interesting, well-integrated, and scholarly."
—Harold T. Christensen, Purdue University

"An excellent coverage of the field with fact and theory judiciously integrated and arranged."
—Joseph G. Keegan, Fordham University

American Book Company

COLLEGE DIVISION • 55 FIFTH AVENUE, NEW YORK 3, N. Y.

PRESENT-DAY PSYCHOLOGY

EDITED BY A. A. ROBACK

A definitive volume of 40 original contributions embracing practically the whole range of psychology from the neurological basis to military and parapsychology, each chapter written by an expert in his field expressly for this work.

FROM THE TABLE OF CONTENTS

RECENT FINDINGS IN GENERAL NEUROLOGY
Joseph G. Keegan

ISSUES AND RESULTS IN SENSORY PSYCHOLOGY
P. Ratoosh

THE STATUS OF EMOTION IN CONTEMPORARY PSYCHOLOGY
Magda Arnold

MIDCENTURY PSYCHOSOMATICS
James W. D. Hartman

CLINICAL PSYCHOLOGY
W. G. Eliasberg

RECENT DEVELOPMENTS IN PSYCHOTHERAPY
Emil A. Gutheil

INDIVIDUAL PSYCHOLOGY
Rudolf Dreikurs

ABNORMAL PSYCHOLOGY
Philip L. Harriman

SOCIAL PSYCHOLOGY
Eugene L. Gaier

APPLIED PSYCHOLOGY
Harold E. Burtt

INTRODUCTION TO EXPERIMENTAL PARAPSYCHOLOGY
J. B. Rhine

SOME RECENT EXPERIMENTAL WORK IN PSYCHODIAGNOSTICS
Werner Wolff

PRESENT-DAY PSYCHOLOGY OF SPEECH
Emil Froeschels

MILESTONES IN PSYCHOANALYSIS
Leon J. Saul and Andrew S. Watson

PROJECTIVE TECHNIQUES IN CONTEMPORARY PSYCHOLOGY
Leopold Bellak

EDUCATIONAL PSYCHOLOGY
Gordon C. Hanson

ADOLESCENCE
Karl C. Garrison

HYPNOTHERAPY
Milton V. Kline

TRENDS IN STATISTICS AND PROBABILITY IN PSYCHOLOGY
Herbert Solomon

INTEGRATIONAL PSYCHOLOGY
Clarence Leuba

Approx. 1,000 pages . . . $12.00

PHILOSOPHICAL LIBRARY, *Publishers*

15 EAST 40TH ST., DESK 185 NEW YORK 16, N. Y.

Expedite shipment by prepayment

YIDDISH SCIENTIFIC INSTITUTE-YIVO

535 West 123rd Street, New York, 27, N.Y.

ANNOUNCEMENT OF THE SEVENTH ANNUAL YIVO CONTEST 1955

The Jewish community in the United States and Canada, now comprising nearly one-half of the Jewish population of the world, has steadily grown in importance. Three hundred years old, it has made its contributions to America and has become a significant factor in the life of the Jewish people the world over.

To stimulate interest in the study of the processes through which this community passes and the problems it faces, the Yiddish Scientific Institute-YIVO announces a contest for original research papers. Following are the terms of the contest:

1. YIVO offers three cash awards of $300, $200 and $100 as first, second and third prizes respectively for the three best studies and essays on subjects related to Jewish life in the United States or Canada.
2. Papers may be on any topic pertaining to Jewish life in America. Each entry should be based on valid sources and should reflect the independent research efforts of the contestant. Entries will be judged for content rather than style.
3. Each entry shall contain about 5,000 words. At least two copies of each entry shall be submitted.
4. Any graduate student or college senior is eligible.
5. All entries remain the property of YIVO, which reserves the right to publish any of them in whole or in part.
6. The closing date for submitting entries is October 15, 1955. Awards will be announced at the YIVO conference in January 1956.

***All manuscripts and other communications should be addressed to the* Commission on Research, YIVO, 535 West 123rd St., New York, 27, N.Y.**

INTERNATIONAL SOCIAL SCIENCE BULLETIN

published quarterly by the United Nations Educational, Scientific and Cultural Organization, 19 Avenue Kléber, Paris 16e.

Just issued: Vol. VI, No. 4

MATHEMATICS AND THE SOCIAL SCIENCES

Methods and Results—Research Centres

Contributors to this issue: CLAUDE LÉVI-STRAUSS, B. DE FINETTI, C. CHERRY, L. FESTINGER, A. TUSTIN, R. C. BOOTON, JR., G. TINTNER, P. THIONET, E. SIBLEY

Review of Documents, Periodicals and Books—News and Announcements

Annual subscription: $3.50
Per copy: $1.00

Send subscriptions to:

U. S. DISTRIBUTOR:

UNESCO PUBLICATIONS SERVICE

475 Fifth Avenue **New York 17, N. Y.**

Journal of Abnormal and Social Psychology

AVAILABLE BACK ISSUES

Year	Volume	Available Numbers						Price per Number	Price per Volume
1906–07	1	—	—	—	—	—	—	—	—
1907–08	2	—	—	3	—	—	6	$2.00	—
1908–09	3	—	—	—	—	—	—	—	—
1909–10	4	—	—	—	—	—	—	—	—
1910–11	5	—	2	—	—	—	6	$2.00	—
1911–12	6	—	—	—	—	—	—	—	—
1912–13	7	—	—	—	—	—	—	—	—
1913–14	8	—	—	—	—	—	—	—	—
1914–15	9	—	—	—	—	—	—	—	—
1915–16	10	—	—	—	—	—	—	—	—
1916–17	11	—	—	—	—	—	—	—	—
1917–18	12	—	—	—	—	—	—	—	—
1918–19	13	1	2	—	—	—	—	$2.00	—
1919–20	14	—	—	3	4	5	6	$2.00	—
1920–21	15	1	—	—	4	—	—	$2.00	—
1921–22	16	—	—	—	—	—	—	—	—
		Apr.	July	Oct.	Jan.				
1922–23	17	1	—	—	—			$2.00	—
1923–24	18	—	—	—	—			—	—
1924–25	19	1	2	3	4			$2.00	$8.00
1925–26	20	—	—	—	—			—	—
1926–27	21	—	—	—	—			—	—
1927–28	22	—	—	—	4			$2.00	—
1928–29	23	—	—	—	4			$2.00	—
1929–30	24	—	2	3	4			$2.00	—
1930–31	25	1	2	3	4			$2.00	$8.00
1931–32	26	1	2	3	4			$2.00	$8.00
1932–33	27	1	2	3	4			$2.00	$8.00
1933–34	28	1	2	3	4			$2.00	$8.00
1934–35	29	1	2	3	4			$2.00	$8.00
1935–36	30	1	2	3	4			$2.00	$8.00
1936–37	31	1	2	3	4			$2.00	$8.00
		Jan.	Apr.	July	Oct.				
1937	32	1	2	3	& 4			$2.00	$8.00
1938	33	1	2	3	4			$2.00	$8.00
1939	34	1	—	3	4			$2.00	—
1940	35	—	—	—	—			—	—
1941	36	—	—	—	—			—	—
1942	37	1	2	3	4			$2.00	$8.00
1943	38	1	2	3	4			$2.00	$8.00
1944	39	1	2	3	4			$2.00	$8.00
1945	40	1	2	3	4			$2.00	$8.00
1946	41	1	2	3	4			$2.00	$8.00
1947	42	1	2	—	4			$2.00	—
1948	43	1	2	—	4			$2.00	—
1949	44	1	2	3	4			$2.00	$8.00
1950	45	1	2	3	4			$2.00	$8.00
1951	46	1	2	3	4			$2.00	$8.00
1952	47	1	2	3	4			$2.00	$8.00
1953	48	1	2	3	4			$2.00	$8.00
1954	49	1	2	3	4			$2.00	$8.00
1955	50 51	By subscription $10.50 (2 volumes)						$2.00	—

Beginning in 1955 there are two volumes a year, three issues in each volume.

Postage prepaid on U. S. orders. Add $.50 per volume on foreign orders. All stock subject to prior sale.

The American Psychological Association gives the following discounts:

10% on orders of $50.00 or over for any one journal
20% on orders of $100.00 or over for any one journal
30% on orders of $150.00 or over for any one journal

Orders for available back issues and for current subscriptions should be addressed to:

American Psychological Association

1333 Sixteenth Street N. W. Washington 6, D. C.

The Dryden Press

JUST PUBLISHED:

SCHOOLS OF PSYCHOANALYTIC THOUGHT

By RUTH L. MUNROE

This long-awaited work by a noted psychologist is at once an exposition, a critique, and an attempt at integration.

The introductory chapters discuss the concepts held in common by all schools. Then, in five chapters, Dr. Munroe presents the theories of Freud and those Freudians who essentially accept the libido theory. The remaining eight chapters describe the positions of several psychoanalysts who have essentially rejected the libido theory.

Strategically interjected throughout the expository and comparative discussions are the author's critical comments (in double-column format). These critiques are of two kinds: reflections of the clinical psychologist on problems of patients, normal people, our culture, and empirical research; and discussions of a philosophical order, leading toward the integration attempted by the author in this book.

The volume has been forged out of reading, out of observation, and out of direct communication with experts to whom the discussion of each school was submitted for criticism, evaluation, and correction. In no case, however, the final work reflects the aid, the interpretation and the documenting responses of these acknowledged leaders in the field.

"The book is written 'from the inside,'" writes Theodore Newcomb, "The author has taken the time and the living room to enter sympathetically into the most subtle aspects of the various viewpoints she discusses. It is not just superficially comprehensive; it is comprehensive with the comparative discussion of the theories and the theory makers."

SIX APPROACHES TO PSYCHOTHERAPY

Edited by JAMES L. McCARY

With the assistance of DANIEL E. SHEER

How can one counteract the tendency of students to develop a too-limited view of psychotherapy? In this symposium, six authorities present a background to and a description of the leading recognized approaches: Client-Centered Psychotherapy (*Nicholas Hobbs*), Hypnotherapy (*Lewis Wolberg*), Group Psychotherapies (*J. R. Slavson*), Psychotherapy Based on Psychoanalytic Principles (*Norman Reider*), Directive and Eclectic Personality Counseling (*Frederick Thorne*), Psychodrama (*J. L. Moreno*).

In the final chapter, Professor Sheer examines the interrelationships of these six techniques and theories and offers a tentative proposal for their integration within a general framework. *List, 5.75.*

31 W. 54th St. THE DRYDEN PRESS New York 19

WILEY

THEORIES OF PERCEPTION AND THE CONCEPT OF STRUCTURE

A Review and Critical Analysis
with an Introduction to a Dynamic-Structural Theory of Behavior

By Floyd H. Allport, Syracuse University. This unique, double-purpose volume provides the first truly exhaustive, organized survey of the current major theories of perception. It also offers a foundation for future systematic work in the area of a general theory of structure in behavior.

The work opens with a review of the essential facts and characteristics of perception, critically examining them in the light of scientific method. It then presents each of the current major theories separately, bringing out their strong points and weaknesses and correlating them with one another. From this, the author develops a synthesis of the major generalizations more or less common to all the theories.

From the examination the need for a new and general theory of structure in perception and behavior studies emerges clearly. To answer it, the author proposes a dynamic-explicit theory of structure. He then re-examines the various theories in the light of his new theory sheds on their problems. 1955. 799 pages. $8.50.

PSYCHOLOGICAL STATISTICS, Second Edition

By Quinn McNemar, Stanford University. Here is a clear and concise treatment of the statistical techniques most frequently used in psychological research. The presentation stresses assumptions and interpretations rather than routine computational procedures. It offers a complete introduction to statistical methods that moves from simple to more advanced topics and includes an exposition of the elementary principles of statistical inference which was a feature of the earlier edition.

New material in the second edition includes: a development of the principles of hypothesis testing by the use of the binomial distribution and the normal curve as an approximation; a method for handling several correlated proportions and the exact probability method for fourfold tables; and a short chapter on nonparametric methods. *One of the Wiley Publications in Psychology, Herbert S. Langfeld, Editor.* 1955. 408 pages. $6.00.

SOME SOVIET PORTRAITS

By Raymond A. Bauer, Massachusetts Institute of Technology, with the collaboration of Edward Wesolik. A group of fictional sketches based on fact, illustrating the patterns of [illegible] and frustration encountered by the "typical" Soviet citizen. A Technology Press book. M.I.T. 1955. 250 pages. $3.95.

Send today for complete descriptive circulars

JOHN WILEY & SONS, Inc. 440 Fourth Ave., New York 16, N.Y.